

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2003, 14:10:22 ; Search time 93 seconds  
(without alignments)  
27.748 Million cell updates/sec

Title: US-09-868-293B-2  
Perfect score: 46  
Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL.23.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID     | Description        |
|------------|-------|-------------|--------|--------|--------------------|
| 1          | 37    | 80.4        | 581    | Q9U1J3 | Q9U1J3 drosophila  |
| 2          | 36    | 78.3        | 2993   | Q8NMS0 | Q8NMS0 corynebacte |
| 3          | 36    | 78.3        | 3022   | Q8FMV7 | Q8FMV7 corynebacte |
| 4          | 35    | 76.1        | 180    | Q8EIS3 | Q8EIS3 shewanella  |
| 5          | 35    | 76.1        | 190    | Q8ZT16 | Q8ZT16 pyrobaculum |
| 6          | 35    | 76.1        | 734    | Q16791 | Q16791 caenorhabdi |
| 7          | 35    | 76.1        | 966    | Q9CAD8 | Q9CAD8 arabidopsis |
| 8          | 34    | 73.9        | 282    | Q52497 | Q52497 raietonia s |
| 9          | 34    | 73.9        | 282    | Q8XRH5 | Q8XRH5 raietonia s |
| 10         | 34    | 73.9        | 398    | Q9KY68 | Q9KY68 streptomyce |
| 11         | 34    | 73.9        | 401    | Q9PDM2 | Q9PDM2 xylella fas |
| 12         | 34    | 73.9        | 427    | Q8XM84 | Q8XM84 clostridium |
| 13         | 34    | 73.9        | 508    | Q82414 | Q82414 papaver som |
| 14         | 34    | 73.9        | 529    | Q99PR5 | Q99PR5 cavia porce |
| 15         | 34    | 73.9        | 584    | Q8PVR8 | Q8PVR8 methanocarc |
| 16         | 34    | 73.9        | 728    | Q9XUQ1 | Q9XUQ1 caenorhabdi |

## ALIGNMENTS

### RESULT 1

|        |  |              |      |         |
|--------|--|--------------|------|---------|
| Q9U1J3 | Q9U1J3   | PRELIMINARY; | PRT; | 581 AA. |
| AC     | Q9U1J3; Q9V7P7;  |              |      |         |
| DT     | 01-MAY-2000 (TREMBLrel. 13, Created)                                   |              |      |         |
| DT     | 01-MAY-2000 (TREMBLrel. 13, Last sequence update)                      |              |      |         |
| DT     | 01-MAR-2003 (TREMBLrel. 23, Last annotation update)                    |              |      |         |
| DE     | Malate dehydrogenase (NADP-dependent oxaloacetate decarboxylating),    |              |      |         |
| DE     | malic enzyme (EC 1.1.1.40) (MENL-1 protein).                           |              |      |         |
| GN     | MENL-1 OR CG7964.  |              |      |         |
| OS     | Drosophila melanogaster (Fruit fly).                                   |              |      |         |
| OC     | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;          |              |      |         |
| OC     | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;             |              |      |         |
| OC     | Ephydroidea; Drosophilidae; Drosophila.                                |              |      |         |
| OX     | NCBI TaxID=7227;   |              |      |         |
| RN     | {1}_   |              |      |         |
| RP     | SEQUENCE FROM N.A.   |              |      |         |
| RC     | STRAIN=OREGON R;   |              |      |         |
| RA     | Farkas R., Danis P., Mechler B.M.;                                     |              |      |         |
| RL     | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.                |              |      |         |
| RN     | {2}_   |              |      |         |
| RP     | SEQUENCE FROM N.A.   |              |      |         |
| RC     | STRAIN=BERKELEY;   |              |      |         |
| RX     | MEDLINE=20196006; PubMed=10731132;                                     |              |      |         |
| RA     | Adams M.D., Celisniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,      |              |      |         |
| RA     | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,      |              |      |         |
| RA     | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,    |              |      |         |
| RA     | Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,          |              |      |         |
| RA     | Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,   |              |      |         |
| RA     | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  |              |      |         |
| RA     | Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  |              |      |         |
| RA     | Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,    |              |      |         |
| RA     | Beeson K.Y., Bertson P.V., Berman B.P., Bhandari D., Bolshakov S.,     |              |      |         |
| RA     | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,         |              |      |         |
| RA     | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  |              |      |         |
| RA     | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,          |              |      |         |
| RA     | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,      |              |      |         |
| RA     | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., |              |      |         |
| RA     | Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., |              |      |         |

Q8ia76 caenorhabdi  
Q8ia75 caenorhabdi  
Q8cnu3 staphylococ  
Q8b9i9 rachiplusia  
Q8ell1 oceanobacil  
Q9et95 synecococc  
Q8dje2 synecococc  
O57253 vaccinia vi  
Q8yvv9 anabaena sp  
Q9a5l9 caulobacter  
Q9xb55 zymomonas m  
Q9q0h9 influenza a  
Q9q0i0 influenza a  
Q8n7d4 homo sapien  
Q8n7e3 homo sapien  
Q9j4a2 influenza a  
Q9ukn1 homo sapien  
Q965w8 caenorhabdi  
Q98ta2 brachydanio  
Q8ayh1 brachydanio  
Q9uxs0 haloarcula  
Q9p7w8 schizosacch  
Q9hms5 halobacteri  
Q9nes7 leishmania  
Q8cdi6 mus musculu  
Q9zkk7 helicobacte  
Q8g1g0 streptococc  
Q9qst4 human immun  
Q905e6 human immun

17 34 73.9 1068 5 Q8IA76  
18 34 73.9 1075 5 Q8IA75  
19 33 71.7 153 16 Q8CNU3  
20 33 71.7 156 12 Q8B9I9  
21 33 71.7 161 16 Q8ELL1  
22 33 71.7 175 2 Q9ET95  
23 33 71.7 175 16 Q8DJE2  
24 33 71.7 179 12 Q57253  
25 33 71.7 206 16 Q8YVV9  
26 33 71.7 257 16 Q9A5L9  
27 33 71.7 276 2 Q9XB55  
28 33 71.7 402 12 Q9Q0H9  
29 33 71.7 402 12 Q9Q0I0  
30 33 71.7 412 4 Q8N7D4  
31 33 71.7 424 4 Q8N7E3  
32 33 71.7 560 12 Q9J4A2  
33 33 71.7 585 4 Q9UKN1  
34 33 71.7 684 5 Q965W8  
35 33 71.7 725 13 Q98TA2  
36 33 71.7 725 13 Q8AYH1  
37 33 71.7 773 1 Q9UXS0  
38 33 71.7 780 3 Q9P7W8  
39 33 71.7 784 17 Q9HMS5  
40 33 71.7 894 5 Q9NE57  
41 33 71.7 1109 11 Q8CDI6  
42 33 71.7 2231 16 Q9ZKK7  
43 32 69.6 74 2 Q8GLG0  
44 32 69.6 80 15 Q9QST4  
45 32 69.6 80 15 Q905E6

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulyov G., Milschina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -I- SIMILARITY: BELONGS TO THE MALIC ENZYMES FAMILY.  
 DR EMBL; AJ251543; CAB64260.1; -;  
 DR EMBL; AE003807; AAF58000.2; -;  
 DR HSPB; P23368; IOR6.  
 DR FLYBase; FBgn0029154; Men1-1.  
 DR InterPro; IPR001891; Malic\_oxred.  
 DR Pfam; PF00390; malic; 1.  
 DR Pfam; PF03949; malic N; 1.  
 DR PRINTS; PR00072; WALOXRDASE.  
 DR PROSITE; PS00331; MALIC\_ENZYMES; 1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 581 AA; 64692 MW; 47D49B1DE58D6F44 CRC64;

Query Match 80.4%; Score 37; DB 5; Length 581;  
 Best Local Similarity 88.9%; Pred. No. 70;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TESLETLVE 10  
 :|:|:|:|:|  
 Db 393 TKSLETLVE 401

RESULT 2  
 Q8NMS0 PRELIMINARY; PRT; 2993 AA.  
 ID Q8NMS0  
 AC Q8NMS0;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DT 3-oxoacyl-(acyl-carrier-protein) synthase (EC 2.3.1.85).  
 GN CGL2495.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005281; BAB99888.1; -;  
 DR InterPro; IPR001227; Ac transferase.  
 DR InterPro; IPR003965; Fatty\_acid\_synth.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR002539; MacC.dehydratase.  
 DR InterPro; IPR001638; SBP bac 3.  
 DR Pfam; PF00698; Acyl transf; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.

DR Pfam; PF01575; MacC.dehydratase; 1.  
 DR PRINTS; PR01483; FASYNTHASE.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR PROSITE; PS01039; SBP\_BACTERIAL\_3; 1.  
 KW Transferase; Acyltransferase; Complete proteome.  
 SQ SEQUENCE 2993 AA; 316347 MW; 14CCEA44FAC3561B CRC64;

Query Match 78.3%; Score 36; DB 16; Length 2993;  
 Best Local Similarity 77.8%; Pred. No. 6e+02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TESLETLVE 10  
 :|:|:|:|:|  
 Db 1740 TDSLETLVE 1748

RESULT 3  
 Q8FMV7 PRELIMINARY; PRT; 3022 AA.  
 ID Q8FMV7  
 AC Q8FMV7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Fatty-acid synthase II (EC 2.3.1.85).  
 GN FASB OR CE2392.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Usuda Y., Sugimoto S.;  
 RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005222; BAC19202.1; -;  
 KW Acyltransferase; Transferase; Complete proteome.  
 SQ SEQUENCE 3022 AA; 319680 MW; 660C346DFCC35335 CRC64;

Query Match 78.3%; Score 36; DB 16; Length 3022;  
 Best Local Similarity 77.8%; Pred. No. 6.1e+02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TESLETLVE 10  
 :|:|:|:|:|  
 Db 1767 TDSLETLVE 1775

RESULT 4  
 Q8EIS3 PRELIMINARY; PRT; 180 AA.  
 ID Q8EIS3  
 AC Q8EIS3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Isochorismatase family protein.  
 GN SO0762.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RA MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Caidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim A., Lee K., Berry K., Lee C.,

RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RL Shewanella oneidensis."; Science 282:2012-2018(1998).  
 RN [2]  
 DR EMBL; A501521; AAN53838.1; -;  
 DR TIGR; SO0762; -;  
 KW Complete proteome.  
 SQ SEQUENCE 180 AA; 20216 MW; 47A7851AEBE80488 CRC64;

Query Match 76.1%; Score 35; DB 16; Length 180;  
 Best Local Similarity 70.0%; Pred. No. 52;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

|||||:|:|

Db 59 TTEELKTLLE 68

RESULT 5

Q82T16 PRELIMINARY; PRT; 190 AA.  
 AC Q82T16;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Indolepyruvate ferredoxin oxidoreductase alpha subunit part 1,  
 DE authentic frameshift.  
 GN PAE3478.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum."; J. Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 RL EMBL; AB009933; AAL64947.1; -;  
 DR InterPro; IPR002880; POR\_N.  
 DR Pfam; PF01855; POR\_N; 1  
 KW Pyruvate; Complete proteome.  
 SQ SEQUENCE 190 AA; 20668 MW; 86149CDC5F1CCA86 CRC64;

Query Match 76.1%; Score 35; DB 17; Length 190;  
 Best Local Similarity 70.0%; Pred. No. 55;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

:|||:|:|

Db 31 STEILETLIE 40

RESULT 6

O16791 PRELIMINARY; PRT; 734 AA.  
 AC O16791;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical 83.9 kDa protein.  
 GN F18A12.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."; Science 282:2012-2018(1998).  
 RN [2]  
 DR EMBL; A501521; AAN53838.1; -;  
 DR TIGR; SO0762; -;  
 KW Complete proteome.  
 SQ SEQUENCE 180 AA; 20216 MW; 47A7851AEBE80488 CRC64;

QY 2 TESLETLVE 10

|||||:|:|

Db 364 TESLTLIE 372

RESULT 7

Q9CAD8 PRELIMINARY; PRT; 966 AA.  
 AC Q9CAD8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Putative disease resistance protein.  
 GN F24D7.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=1130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana." Nature 408:816-820(2000).  
 RL Nature 408:816-820(2000).  
 DR EMBL; AC011622; AAG52419.1; -;

Mol. Plant Microbe Interact. 5:384-389(1992).

[5]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=93302711; PubMed=8316211;

Gough C.L., Genin S., Lopes V., Boucher C.A.;

"Homology between the HrpO protein of Pseudomonas solanacearum and bacterial proteins implicated in a signal peptide-independent secretion mechanism.";

Mol. Gen. Genet. 239:378-392(1993).

[6]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=96347139; PubMed=8736546;

Bogdanove A.J., Beer S.V., Bonas U., Boucher C.A., Collmer A., Coplin D.L., Cornelis G.R., Huang H.C., Hutcheson S.W., Panopoulos N.J., Van Gijsegem F.;

"Unified nomenclature for broadly conserved hrp genes of phytopathogenic bacteria.";

Mol. Microbiol. 20:681-683(1996).

[7]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=94148001; PubMed=8313899;

Arlat M., Van Gijsegem F., Huet J., Pernollet J., Boucher C.A.;

"PopA, a protein which induces a hypersensitivity-like response on specific petunia genotypes, is secreted via the Hrp pathway of Pseudomonas solanacearum.";

EMBO J. 13:543-553(1994).

[8]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=94247349; PubMed=8190064;

Genin S., Boucher C.A.;

"A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of N-terminal domain.";

Mol. Gen. Genet. 243:112-118(1994).

DR EMBL; AJ245811; CAB58259.1; -

DR InterPro; IPR002010; Bac export\_1.

DR InterPro; IPR006304; SpaR\_YscT\_1.

DR Pfam; PF01311; Bac export\_1; 1.

DR PRINTS; PR00953; TYPE3IMRPROT.

DR TIGRFAMs; TIGR01401; flir\_like\_III; 1.

SQ SEQUENCE 282 AA; 29701 MW; 031B93B53DF6337C CRC64;

Query Match 73.9%; Score 34; DB 2; Length 282;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps

QY 1 TTSESLETLV 9

|:|||||:

Db 14 TSSESLETLL 22

RESULT 9

QBXRHS PRELIMINARY; PRT; 282 AA.

AC QBXRHS;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE HRP conserved HRCT transmembrane protein.

GN HRCT OR RSP0872 OR RS01643.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

OG Plasmid megaplasmid.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

OC Ralstoniaceae; Ralstonia.

OX NCBI\_TaxID=305;

RN [1] \_TaxID=305;

RP SEQUENCE FROM N.A.

RX MEDLINE=93113006; PubMed=1472716;

RA Gough C.L., Genin S., Zischek C., Boucher C.A.;

"hrp genes of Pseudomonas solanacearum are homologous to pathogenicity determinants of animal pathogenic bacteria and are conserved among plant pathogenic bacteria.";

Mol. Microbiol. 6:3065-3076(1992).

[4]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=93113006; PubMed=1472716;

Gough C.L., Genin S., Zischek C., Boucher C.A.;

"hrp genes of Pseudomonas solanacearum are homologous to pathogenicity determinants of animal pathogenic bacteria and are conserved among plant pathogenic bacteria.";

Mol. Microbiol. 15:1095-1114(1995).

[3]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=95349395; PubMed=7623665;

Van Gijsegem F., Gough C.L., Zischek C., Genin S., Niqueux E., Arlat M., Barberis P., German S., Castello P., Boucher C.A.;

"The hrp gene locus of Pseudomonas solanacearum which controls a type III secretion system, encodes eight proteins related to components of the flagellar biogenesis complex.";

Mol. Microbiol. 15:1095-1114(1995).

[3]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).

[2]

SEQUENCE FROM N.A.

STRAIN=GMI1000;

MEDLINE=20253307; PubMed=10792715;

Gueneron M., Timmers A.C.J., Boucher C., Arlat M.;

"Two novel proteins, PopB, which has functional nuclear localization RT signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the Hrp-secretion apparatus of Ralstonia solanacearum.";

Mol. Microbiol. 36:261-277(2000).



RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646081; CAD18023.1; --  
 DR InterPro; IPR002010; Bac\_export\_1.  
 DR Pfam; PF01311; Bac\_export\_1; 1.  
 DR PRINTS; PR00953; TYPE3IMPRROT.  
 DR TIGRFAMs; TIGR01401; flir like\_III; 1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 282 AA; 29739 MW; A57915018B221DBA CRC64;

Query Match 73.9%; Score 34; DB 16; Length 282;  
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTSETLTV 9  
 Db 14 TSSETLTL 22

RESULT 10  
 O9KY68 PRELIMINARY; PRT; 398 AA.  
 AC O9KY68;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative NLP/P60 family secreted protein.  
 GN SC04796 OR SC063A.07C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denaparte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";

RL Nature 417:141-147(2002).  
 DR EMBL; AL939121; CAB92659.1; --  
 DR InterPro; IPR000064; NLP\_P60.  
 DR Pfam; PF00877; NLP\_P60; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 398 AA; 42864 MW; 7E1A6BB1C5977B46 CRC64;  
 Query Match 73.9%; Score 34; DB 16; Length 398;  
 Best Local Similarity 77.8%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 TSSETLTV 10  
 Db 169 TSSETLTD 177

RESULT 11  
 Q9PDM2 PRELIMINARY; PRT; 401 AA.  
 AC Q9PDM2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 8-amino-7-oxononanoate synthase.  
 GN XF1357.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5c;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terezzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003967; AAF84166.1; --  
 DR HSP; P12998; IBSO.  
 DR InterPro; IPR003408; Ala\_synthase.  
 DR InterPro; IPR004839; Aminotransf1/2.  
 DR InterPro; IPR001917; NHtransf 2.  
 DR Pfam; PF02490; Ala\_synthase; 1.  
 DR Pfam; PF00155; aminotran1\_2; 1.  
 DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 401 AA; 43861 MW; 6E5F0A94EB3D0C67 CRC64;

Query Match 73.9%; Score 34; DB 16; Length 401;

Best Local Similarity 60.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLTVLVE 10  
|||:|:|:  
Db 374 TTEQIDTLVQ 383

## RESULT 12

O8XM84 Q8XM84 PRELIMINARY; PRT; 427 AA.  
AC O8XM84;  
DT 01-MAR-2002 (TREMREL. 20, Created)  
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)  
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)  
DE Hypothetical protein CPE0806.  
GN CPE0806.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / Type A;  
RX PubMed=11792842;  
RA Shimizu T., Ohnani K., Hirakawa H., Ohshima K., Yamashita A.,  
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003188; BAB80512.1; -;  
DR InterPro; IPR002453; Beta\_tubulin.  
DR PROSITE; PS00228; TUBULIN\_B\_AUTOREG; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 427 AA; 45873 MW; 1CCG77BF6B93F5B9 CRC64;

Query Match 73.9%; Score 34; DB 16; Length 427;  
Best Local Similarity 60.0%; Pred. No. 2e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLTVLVE 10  
|||:|:|:  
Db 396 TTSETLTVLQ 405

## RESULT 13

O82414 O82414 PRELIMINARY; PRT; 508 AA.  
AC O82414;  
DT 01-NOV-1998 (TREMREL. 08, Created)  
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)  
DT 01-OCT-2002 (TREMREL. 22, Last annotation update)  
DE Tyrosine/dopa decarboxylase (EC 4.1.1.28).  
GN TYDC8.  
OS Papaver somniferum (Opium poppy).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
OC Papaveraceae; Papaver.  
OX NCBI\_TaxID=3469;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Marianne;  
RX MEDLINE=98404235; PubMed=9733527;  
RA Faccini P.J., Penzes-Yost C., Samanani N., Kowalchuk B.;  
RT "Expression patterns conferred by tyrosine/dihydroxyphenylalanine  
RT decarboxylase promoters from opium poppy are conserved in transgenic  
RT tobacco.";  
RL Plant Physiol. 118:69-81(1998).  
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -|- TYRDC.  
DR EMBL; AF025432; AAC61841.1; -;  
DR InterPro; IPR002129; Pyridoxal\_dec.

DR Pfam; PF00282; pyridoxal\_dec; 1.  
DR PRINTS; PRO0800; YHDCRBOXLASE.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; 1.  
KW Decarboxylase; Lyase; Pyridoxal phosphate.  
SQ SEQUENCE 508 AA; 56156 MW; 320D2128PB0E59E5 CRC64;

Query Match 73.9%; Score 34; DB 10; Length 508;  
Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTSETLTVLVE 10  
|||:|:|:  
Db 69 TTSETLTVLVE 77

## RESULT 14

Q99PR5 Q99PR5 PRELIMINARY; PRT; 529 AA.  
AC Q99PR5;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)  
DE Enterophilin-1.  
GN 39D1.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Crl.(BFA) BR;  
RA Gassama A., Hullin-Matsuda F., Li R.Y., Nause M., Ragab A.,  
Delagebeaudet C., Simon M.F., Fauvel J., Chap H.;  
RT "Enterophilins, a new family of leucine zipper proteins bearing a  
RT B30.2 domain and associated with enterocyte differentiation.";  
RL J. Biol. Chem. 0:0-0(2001).  
DR EMBL; AF126833; AAK02016.1; -;  
DR InterPro; IPR001870; B302.  
DR InterPro; IPR006574; PRY.  
DR InterPro; IPR003877; SPRY\_receptor.  
DR Pfam; PF00622; SPRY; 1.  
DR SMART; SM00589; PRY; 1.  
DR SMART; SM00449; SPRY; 1.  
SQ SEQUENCE 529 AA; 64503 MW; CEF95AC72546461 CRC64;

Query Match 73.9%; Score 34; DB 11; Length 529;  
Best Local Similarity 87.5%; Pred. No. 2.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTSETLTVLVE 8  
|||:|:|:  
Db 91 TTSETLTVLVE 98

## RESULT 15

Q8PVR8 Q8PVR8 PRELIMINARY; PRT; 584 AA.  
AC Q8PVR8;  
DT 01-OCT-2002 (TREMREL. 22, Created)  
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)  
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)  
DE DNA polymerase beta (EC 2.7.7.7).  
GN MM1894.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goel / GO1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=12125924;  
RA Deppenmeier U., Johann A., Hattsch T., Merkl R., Schmitz R.A.,  
Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,

RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-J., Gottschalk G.,  
RT "The genome of *Mechanosarcina mazei*: evidence for lateral gene  
transfer between Bacteria and Archaea";  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
DR EMBL: AS013426; AM31590.1; -.  
DR InterPro; IPR002054; DNA\_PoIX.  
DR InterPro; IPR004013; PHP\_C.  
DR InterPro; IPR003141; PHP\_N.  
DR Pfam; PF02811; PHP\_C; 1.  
DR Pfam; PF02231; PHP\_N; 1.  
DR PRINTS; PR00869; DNAPOLX.  
KW Transferase; Nucleotidyltransferase; Complete proteome.  
SQ SEQUENCE 584 AA; 65558 MW; 2D404F30D85BD35F CRC64;

Query Match 73.9%; Score 34; DB 17; Length 584;  
Best Local Similarity 87.5%; Pred. No. 2.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
|||||:  
Db 363 ESLETLVE 370

RESULT 16  
Q9XUQ1 PRELIMINARY; PRT; 728 AA.  
AC Q9XUQ1  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DE T08G5.2 protein.  
GN T08G5.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1] TTESLTVE 10  
RP SEQUENCE FROM N.A.  
RA Smye R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z81589; CAB04724.1; -.  
DR WormPep; T08G5.2; CE21175.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 728 AA; 83098 MW; 55EB45F41C2E43C8 CRC64;

Query Match 73.9%; Score 34; DB 5; Length 728;  
Best Local Similarity 70.0%; Pred. No. 3.5e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLTVE 10  
|||||:  
Db 416 TTESLTVE 425

RESULT 17  
Q8IA76 PRELIMINARY; PRT; 1068 AA.  
ID Q8IA76  
AC Q8IA76

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein K10F12.3b.  
GN K10F12.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1] TTESLTVE 10  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RC Wohldmann P., Beck C.;  
RT "The sequence of C. elegans cosmid K10F12.";  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

Query Match 73.9%; Score 34; DB 5; Length 1068;  
Best Local Similarity 60.0%; Pred. No. 5.2e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLTVE 10  
|||||:  
Db 309 TTESLTVE 318

RESULT 18  
Q8IA76 PRELIMINARY; PRT; 1075 AA.  
ID Q8IA76  
AC Q8IA76  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Hypothetical protein K10F12.3a.  
GN K10F12.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1] TTESLTVE 10  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RC Wohldmann P., Beck C.;  
RT "The sequence of C. elegans cosmid K10F12.";  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AF025462; AAN72422.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1075 AA; 121112 MW; 61021111ABF78AFD4 CRC64;

Query Match 73.9%; Score 34; DB 5; Length 1075;
Best Local Similarity 60.0%; Pred. No. 5.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
   ||::|||:|
Db 316 TTDNCETLIE 325

RESULT 19
Q8CNU3 PRELIMINARY; PRT; 153 AA.
AC Q8CNU3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase.
GN SR1438.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016748; AAO05037.1; -.
KW Complete proteome.
SQ SEQUENCE 153 AA; 16458 MW; 39031B85979F4D92 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 153;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
   ||||:|
Db 116 TTESIEQAVE 125

RESULT 20
Q8B9I9 PRELIMINARY; PRT; 156 AA.
AC Q8B9I9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Rachiplusia ou multiple nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=80386;
RN [1]
RP SEQUENCE FROM N.A.
RA Bonning B.C., Harrison R.L.;
RT "The Rachiplusia ou multiple nucleopolyhedrovirus genome sequence.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY145471; AAN28115.1; -.
KW Hypothetical protein.
SQ SEQUENCE 156 AA; 18559 MW; 7BACD77E081DF912 CRC64;

Query Match 71.7%; Score 33; DB 12; Length 156;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
   ||||:|
Db 56 TAESLNTLVD 65

RESULT 21
Q8ELL1 PRELIMINARY; PRT; 161 AA.
AC Q8ELL1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Riboflavin synthase beta chain (EC 2.5.1.9).
GN OB3213.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RA MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP04604; BAC15169.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 161 AA; 17523 MW; 5314E84D339B38F2 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 161;
Best Local Similarity 70.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
   ||||:|
Db 117 TTESIEQAVE 126

RESULT 22
Q9ET95 PRELIMINARY; PRT; 175 AA.
AC Q9ET95;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytochrome c550 like protein.
GN PSBV2.
OS Synechococcus vulcanus (Thermosynechococcus vulcanus), and
OS Thermosynechococcus elongatus.
OC Bacteria; Cyanobacteria; Chroococcales; Thermosynechococcus.
OX NCBI_TaxID=32053, 146786;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.vulcanus, and T.elongatus;
RA Katoh H., Itoh S., Shen J., Ikeuchi M.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.vulcanus, and T.elongatus;
RA Katoh H., Itoh S., Shen J., Ikeuchi M.;
RT "CLONING OF THE GENES FOR CYTOCHROME C550 AND A C550-LIKE PROTEIN FROM
RT THE THERMOPHILIC CYANOBACTERIUM Synechococcus elongatus.";
RL Plant Cell Physiol. 40:89-89(1999).
DR EMBL; AB052598; BAB20064.1; -.
DR EMBL; AB052597; BAB20060.1; -.
DR HSSP; P82603; IFIC.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR003218; Cyt C550.
DR ProDom; PD010841; Cyt c550.1.
DR PROSITE; PS00190; CYTOCHROME C.1.
SQ SEQUENCE 175 AA; 19118 MW; FB95EE3348651C44 CRC64;

Query Match 71.7%; Score 33; DB 2; Length 175;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 TTSELETL 8  
||| |||  
Db 144 TTEQLETL 151

RESULT 23  
ID Q8DJE2 PRELIMINARY; PRT; 175 AA.  
AC Q8DJE2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cytochrome c550 like protein.  
GN TLL1284.  
OS Synechococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP-1;  
RX MEDLINE=22225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1.";  
RL DNA Res. 9:123-130(2002).  
DR EMBL; AF005373; BAC08836.1; -.  
KW Complete proteome.  
SQ SEQUENCE 175 AA; 19118 MW; FB95EE3348651C44 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 175;  
Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSELETL 8  
||| |||  
Db 144 TTEQLETL 151

RESULT 24  
ID O57253 PRELIMINARY; PRT; 179 AA.  
AC O57253;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Putative 20.9k protein.  
GN MVA170R.  
OS Vaccinia virus (strain Ankara).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
OX NCBI\_TaxID=126794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ankara;  
RA Antoine G., Scheiflinger F., Falkner F.G., Dörner F.;  
RT "The complete genomic sequence of the Modified Vaccinia Ankara (MVA)  
RT strain.";  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U94848; AAB96546.1; -.  
SQ SEQUENCE 179 AA; 28665 MW; BCED343C1D1B637F CRC64;

Query Match 71.7%; Score 33; DB 12; Length 179;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTESLETL 10  
||| |||  
Db 34 TESLETL 42

RESULT 25  
ID Q8YV9 PRELIMINARY; PRT; 206 AA.  
AC Q8YV9;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Urease accessory protein F.  
GN ALR0734.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003583; BAB72691.1; -.  
DR InterPro; IPR002639; Uref.  
DR Pfam; PF01730; Uref; 1.  
KW Complete proteome.  
SQ SEQUENCE 206 AA; 22859 MW; 79986D3AF69CF469 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 206;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTESLETL 10  
: || ||| |||  
Db 7 SEGLETL 15

RESULT 26  
ID Q9ASL9 PRELIMINARY; PRT; 257 AA.  
AC Q9ASL9;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Polysaccharide deacetylase.  
GN CC2428.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblym T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Fotocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005912; AAK24399.1; -.  
DR TIGR; CC2428; -.  
DR InterPro; IPR002509; Polysac\_deacet.  
DR Pfam; PF01522; Polysacc\_deacet; 1.  
KW Complete proteome.  
SQ SEQUENCE 257 AA; 27868 MW; 3B39FA0EBDC49A34 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 257;  
Best Local Similarity 60.0%; Pred. No. 1.9e+02;

RT donors of the 'internal' genes of H5N1 viruses in Hong Kong?";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367 (1999).  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; AF156385; AAD48995.1; -.  
DR InterPro; IPR001364; Hemagglutn.  
DR Pfam; PF00509; Hemagglutinin; 1.  
DR PRINTS; PR00329; HEMAGGLUTIN12.  
DR ProDom; PD000225; Hemagglutn; 1.  
DR Envelope protein; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 >402  
FT NON TER 402 402  
FT SEQUENCE 402 AA; 44583 MW; B3C3F5C7BA4808FA CRC64;  
SO

```
Query Match 71.7%; .Score 33; DB 12; Length 402;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 6; .Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```

1 TTESLETLVE 10
:|:|:|:|
30 STEIWDTLVE 39

```

RESULT 29  
90010  
PRELIMINARY: 402 AA.  
90010

| RESULT | 29 | 90010   | PRELIMINARY;                            | PRT; | 402 AA. |
|--------|----|---|---|------|---------|
| D      |    | Q90010  |   |      |         |
| I      |    | Q90010;   |   |      |         |
| T      |    | 01-MAY-2000   | (TREMBLrel. 13, Created)                |      |         |
| T      |    | 01-MAY-2000   | (TREMBLrel. 13, Last sequence update)   |      |         |
| T      |    | 01-MAY-2000   | (TREMBLrel. 22, Last annotation update) |      |         |
| T      |    | 01-OCT-2002   | (TREMBLrel. 22, Last annotation update) |      |         |
| E      |    | Hemagglutinin precursor (Fragment)                        |   |      |         |
| S      |    | Influenza A virus (A/Chicken/Korea/38349-p96323/96(H5N2)) |   |      |         |
| S      |    | Viruses; ssRNA negative-strand viruses; Orthomyxoviridae; |   |      |         |
| C      |    | Influenza A viruses; Influenzavirus A.                    |   |      |         |
| X      |    | NCBI TaxID=97358;   |   |      |         |

|   |
|---|
| RESULT 29   |
| 9Q0I0   |
| D PRELIMINARY; PRT; 402 AA.                                   |
| C Q90I0;  |
| T 01-MAY-2000 (TRENBLrel. 13, Created)                        |
| T 01-MAY-2000 (TRENBLrel. 13, Last sequence update)           |
| T 01-OCT-2002 (TRENBLrel. 22, Last annotation update)         |
| E Hemagglutinin precursor (Fragment).                         |
| S Influenza A virus (A/Chicken/Korea/38349-p36323/96 (H9N2)). |
| S Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;   |
| C Influenza A viruses; Influenzavirus A.                      |

|   |
|---|
| RESULT 29   |
| 9Q0I0   |
| D PRELIMINARY; PRT; 402 AA.                                   |
| C Q90I0;  |
| T 01-MAY-2000 (TRENBLrel. 13, Created)                        |
| T 01-MAY-2000 (TRENBLrel. 13, Last sequence update)           |
| T 01-OCT-2002 (TRENBLrel. 22, Last annotation update)         |
| E Hemagglutinin precursor (Fragment).                         |
| S Influenza A virus (A/Chicken/Korea/38349-p36323/96 (H2N2)). |
| S Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;   |
| C Influenza A viruses; Influenzavirus A.                      |

```

Q8N7D4
ID Q8N7D4 PRELIMINARY; PRT; 412 AA.
AC Q8N7D4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ25770.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshina A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie K., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK098636; BAC05359.1; -.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 46838 MW; 24CC67AD2BFA2AAD CRC64;

Query Match 71.7%; Score 33; DB 4; Length 412;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10
Db 373 ESLETLVE 380
||||:||||

RESULT 31
Q8N7E3 PRELIMINARY; PRT; 424 AA.
ID Q8N7E3;
AC Q8N7E3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ25734.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshina A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie K., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK098600; BAC05347.1; -.
KW Hypothetical protein.
SQ SEQUENCE 424 AA; 47280 MW; C5A461E71ACEBB96 CRC64;

Query Match 71.7%; Score 33; DB 4; Length 424;
Best Local Similarity 87.5%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10
Db 385 ESLETLVE 392
||||:||||

RESULT 32
Q9J4A2 PRELIMINARY; PRT; 560 AA.
ID Q9J4A2
AC Q9J4A2;

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hemagglutinin.
GN H9.
OS Influenza A virus (A/ck/Korea/ms96/96).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
RN [1]_TaxID=123774;
RX NCBI_TaxID=123774;
RP SEQUENCE FROM N.A.
RC STRAIN=a/ck/korea/ms96/96;
RA Lee C.-W., Song C.-S., Lee Y.-J., Mo I.-P., Garcia M., Suarez D.L.,
RA Kim S.-J.;
RT "Molecular and pathogenic characterization of Korean isolates of H9N2
RT avian influenza virus.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: AF203008; AAF69255.1; -.
DR HSSP: P03437; IHTW.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 560 AA; 62673 MW; B935D9D3C8F1114E CRC64;

Query Match 71.7%; Score 33; DB 12; Length 560;
Best Local Similarity 60.0%; Pred. No. 4.2e+02;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
Db 30 STETVDTLVE 39
||||:||||

RESULT 33
Q9UKN1 PRELIMINARY; PRT; 585 AA.
ID Q9UKN1;
AC Q9UKN1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transmembrane mucin 12 (Fragment).
GN MUC12.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon mucosa;
RX MEDLINE=99391252; PubMed=10463611;
RA Williams S.J., McGuckin M.A., Gotley D.C., Eyre H.J., Sutherland G.R.,
RA Antalis T.M.;
RT "Two novel mucin genes down-regulated in colorectal cancer identified
RT by differential display.";
RL Cancer Res. 59:4083-4089(1999).
DR EMBL: AF147790; AAD55678.1; -.
DR Genew; HGNC:7510; MUC12.
DR InterPro: IPR006209; EGF_like.
DR Pfam: PF01390; SEA; 1.
DR SMART: SM00200; SEA; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
FT NON_TER 1
SQ SEQUENCE 585 AA; 64015 MW; F3F063E930058DB4 CRC64;

```

Query Match 71.7%; Score 33; DB 4; Length 585;  
Best Local Similarity 87.5%; Pred. No. 4.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTSELETL 8  
DB 226 STSELETL 233

RESULT 34  
Q965W8 PRELIMINARY; PRT; 684 AA.  
AC Q965W8;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Hypochemical protein Y39H10A.6.  
GN Y39H10A.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Leonard S.;  
RT "The sequence of C. elegans cosmid Y39H10A.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024770; AAK84605.2; -;  
KW Hypothetical protein.  
SQ SEQUENCE 684 AA; 77978 MW; 879FD1056BAEDB4B CRC64;

Query Match 71.7%; Score 33; DB 5; Length 684;  
Best Local Similarity 77.8%; Pred. No. 5.2e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TSELETL 10  
DB 71 TSSVETL 79

RESULT 35  
Q98TA2 PRELIMINARY; PRT; 725 AA.  
AC Q98TA2;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE 17-beta-hydroxysteroid dehydrogenase type 4.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Marjanovic Z., Breitling R., Perovic D., Moeller G., Adamski J.;  
RT "Cloning and characterization of 17 beta-hydroxysteroid dehydrogenase  
RT in zebrafish.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC (SDR) FAMILY.  
DR EMBL; AF241285; AAK27967.1; -;  
DR HSSP; O70351; IE6W.  
DR InterPro; IPR002198; ADH short.  
DR InterPro; IPR002539; MacC\_dehydratase.  
DR InterPro; IPR003033; SCP2.  
DR Pfam; PF00106; adh\_short; 1.  
DR Pfam; PF01575; MacC\_dehydratase; 1.  
DR Pfam; PF02036; SCP2; 1.  
DR PRINTS; PR00080; SDRFAMILY.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
KW Oxidoreductase.  
SQ SEQUENCE 725 AA; 78596 MW; 7BE76087293771F7 CRC64;

Query Match 71.7%; Score 33; DB 13; Length 725;  
Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
DB 290 ESLQTLVE 297

RESULT 36  
Q8AYH1 PRELIMINARY; PRT; 725 AA.  
AC Q8AYH1;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE 17-beta-hydroxysteroid dehydrogenase type 4.  
GN HSD17B4.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mindnich R., Adamski J.;  
RT "Structure of hsd17b4 gene in zebrafish.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF439319; AAN62014.1; -;  
DR EMBL; AF439315; AAN62014.1; JOINED.  
DR EMBL; AF439316; AAN62014.1; JOINED.  
DR EMBL; AF439317; AAN62014.1; JOINED.  
DR EMBL; AF439318; AAN62014.1; JOINED.  
SQ SEQUENCE 725 AA; 78624 MW; E6566A078130AE8A CRC64;

Query Match 71.7%; Score 33; DB 13; Length 725;  
Best Local Similarity 87.5%; Pred. No. 5.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
DB 290 ESLQTLVE 297

RESULT 37  
Q9UXS0 PRELIMINARY; PRT; 773 AA.  
ID Q9UXS0  
AC Q9UXS0;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE TrII protein.  
GN TrII.  
OS Haloarcula vallismortis.  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Haloarcula.  
OX NCBI\_TaxID=28442;  
RN [1]



```
RP SEQUENCE FROM N.A.
RC STRAIN=SP1;
RA Rodewald K., Seidel R., Engelhard M., Oesterhelt D.;
RT "Primary structure of vHtrII, a transducer protein from Haloarcula
RT vallismortis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ249640; CAB56464.1; -.
DR HSSP; P02942; 1Q07.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPeignial; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
SQ SEQUENCE 773 AA; 82592 MW; 9CFBFF01183AC6A CRC64;

Query Match 71.7%; Score 33; DB 1; Length 773;
Best Local Similarity 60.0%; Pred. No. 5.9e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLTVE 10
DB 663 TVDALETIVE 672

RESULT 38
Q9PW78 SEQUENCE FROM N.A.
ID Q9PW78 PRELIMINARY; PRT; 780 AA.
AC Q9PW78;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 88.0 kDa protein.
GN SPBC1703.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Cadieu E., Lelaure V.,
RA Gallibert F.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL136536; CAB66446.1; -.
DR GeneDB SPombe; SPBC1703.02; -.
DR InterPro; IPR001606; ARID.
DR Pfam; PF01388; ARID; 1.
DR SMART; SM00501; BRIGHT; 1.
KW Hypothetical protein.
SQ SEQUENCE 780 AA; 87994 MW; 76416645839716BC CRC64;

Query Match 71.7%; Score 33; DB 3; Length 780;
Best Local Similarity 87.5%; Pred. No. 5.9e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10
DB 695 DSLETLVE 702

RESULT 39
Q9HWM5 SEQUENCE FROM N.A.
ID Q9HWM5 PRELIMINARY; PRT; 784 AA.
AC Q9HWM5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent RNA helicase homolog eIF-4A.
SQ SEQUENCE 784 AA; 85599 MW; 3D52762274161B90 CRC64;

Query Match 71.7%; Score 33; DB 17; Length 784;
Best Local Similarity 60.0%; Pred. No. 6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTSETLTVE 10
DB 689 TTETMETIAE 698

RESULT 40
Q9NES7 SEQUENCE FROM N.A.
ID Q9NES7 PRELIMINARY; PRT; 894 AA.
AC Q9NES7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Conserved hypothetical protein L5213T.05.
GN L5213T.05.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Zimmermann W., Wambutt R., Ivens A.C., Quail M., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RX Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RX Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
RL EMBL; AL352992; CAB88226.1; -.
DR InterPro; IPR002641; Patatin.
```

DR Pfam: PF01734; Patatin; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 894 AA; 96609 MW; C6FE1737E63ED32E CRC64;  
 Query Match 71.7%; Score 33; DB 5; Length 894;  
 Best Local Similarity 70.0%; Pred. No. 6.8e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TTESLETIVE 10  
 ||||:||  
 Db 132 TTESYKTFVE 141

Search completed: October 30, 2003, 14:17:57  
 Job time : 97 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2003, 14:03:06 ; Search time 23 seconds  
(without alignments)  
20.446 Million cell updates/sec

Title: US-09-868-293B-2  
Perfect score: 46  
Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description        |
|------------|-------|-------------|--------|--------------|--------------------|
| 1          | 46    | 100.0       | 128    | 1 RL7_CHLPN  | Q929a1 chlamydia p |
| 2          | 46    | 100.0       | 129    | 1 RL7_CHLMU  | P38001 chlamydia m |
| 3          | 46    | 100.0       | 129    | 1 RL7_CHLTR  | O84318 chlamydia t |
| 4          | 38    | 82.6        | 946    | 1 AMVG_CANAL | O74254 candida alb |
| 5          | 36    | 78.3        | 316    | 1 ML34_ARATH | Q98sk7 arabidopsis |
| 6          | 34    | 73.9        | 438    | 1 TOLC_VIBCH | Q9kzy1 vibrio chol |
| 7          | 33    | 71.7        | 124    | 1 VB03_VACCC | P21000 vaccinia vi |
| 8          | 33    | 71.7        | 154    | 1 RISE_STAAM | Q93in8 staphylococ |
| 9          | 33    | 71.7        | 155    | 1 Y063_NPVAC | P41466 autographa  |
| 10         | 33    | 71.7        | 167    | 1 VB03_VACCV | Q01226 vaccinia vi |
| 11         | 33    | 71.7        | 433    | 1 HTR2_HALVA | P42258 haloarcula  |
| 12         | 33    | 71.7        | 3386   | 1 POLG_DEN4  | P09866 d genome po |
| 13         | 33    | 71.7        | 3924   | 1 ANK2_HUMAN | Q01484 homo sapien |
| 14         | 33    | 71.7        | 4447   | 1 PKSK_BACSU | P40803 bacillus su |
| 15         | 32    | 69.6        | 45     | 1 VPU_HV123  | P08805 human immu  |
| 16         | 32    | 69.6        | 156    | 1 RISE_ECOLI | P25540 escherichia |
| 17         | 32    | 69.6        | 306    | 1 UCPI_MOUSE | P12242 mus musculu |
| 18         | 32    | 69.6        | 306    | 1 UCPI_RAT   | P04633 rattus norv |
| 19         | 32    | 69.6        | 406    | 1 ISDF_HELPY | O25664 h ispd/ispf |
| 20         | 32    | 69.6        | 409    | 1 ISDF_HELPJ | Q9zm19 h ispd/ispf |
| 21         | 32    | 69.6        | 422    | 1 MURD_HELPJ | Q9zly0 helicobacte |
| 22         | 32    | 69.6        | 598    | 1 YK21_SCHPO | O13910 schizosacch |
| 23         | 32    | 69.6        | 1025   | 1 DM3_YEAST  | Q12176 saccharomyc |
| 24         | 32    | 69.6        | 1147   | 1 TEAL_SCHPO | P87061 schizosacch |
| 25         | 32    | 69.6        | 1338   | 1 ACIN_MOUSE | Q9jix8 mus musculu |
| 26         | 32    | 69.6        | 1341   | 1 ACIN_HUMAN | Q9ukv3 homo sapien |
| 27         | 32    | 69.6        | 1375   | 1 RPOB_COXBU | O87903 coxiella bu |
| 28         | 31    | 67.4        | 141    | 1 MRAZ_MYCGE | P47463 mycoplasma  |
| 29         | 31    | 67.4        | 242    | 1 FABG_HAEIN | P43713 haemophilus |
| 30         | 31    | 67.4        | 367    | 1 Y638_RHILO | Q98mcl rhizobium l |
| 31         | 31    | 67.4        | 510    | 1 DDC_ACIBA  | Q34308 acinetobact |
| 32         | 31    | 67.4        | 517    | 1 SEST_CAEL  | Q9n4d6 caenorhabdi |
| 33         | 31    | 67.4        | 577    | 1 YG5U_YEAST | P53333 saccharomyc |

|           |  |           |      |     |     |            |        |             |  |
|-----------|--|-----------|------|-----|-----|------------|--------|-------------|--|
| RESULT 1  |  |           |      |     |     |            |        |             |  |
| RL7_CHLPN | 34   | 31        | 67.4 | 595 | 1   | PRIM_CHLTR | 084799 | chlamydia t |  |
| ID        | RL7_CHLPN  | STANDARD; | PRT; | 128 | AA. |            |        |             |  |
| AC        | Q929A1: Q9JQ70;  |           |      |     |     |            |        |             |  |
| DT        | 30-MAY-2000 (Rel. 39, Created)   |           |      |     |     |            |        |             |  |
| DT        | 30-MAY-2000 (Rel. 39, Last sequence update)  |           |      |     |     |            |        |             |  |
| DT        | 28-FEB-2003 (Rel. 41, Last annotation update)  |           |      |     |     |            |        |             |  |
| DE        | 50S ribosomal protein L7/L12   |           |      |     |     |            |        |             |  |
| GN        | RPLL OR RL7 OR CPN0080 OR CP0695.  |           |      |     |     |            |        |             |  |
| OS        | Chlamydia pneumoniae (Chlamydophila pneumoniae)  |           |      |     |     |            |        |             |  |
| OC        | Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.  |           |      |     |     |            |        |             |  |
| OX        | NCBI_TaxID=83358;  |           |      |     |     |            |        |             |  |
| RN        | [1]  |           |      |     |     |            |        |             |  |
| RP        | SEQUENCE FROM N.A.   |           |      |     |     |            |        |             |  |
| RC        | STRAIN=CWL029;   |           |      |     |     |            |        |             |  |
| RX        | MEDLINE=99206606; PubMed=10192388;   |           |      |     |     |            |        |             |  |
| RA        | Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,   |           |      |     |     |            |        |             |  |
| RA        | Ollinger L., Grimwood J., Davis R.W., Stephens R.S.;   |           |      |     |     |            |        |             |  |
| RT        | "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";   |           |      |     |     |            |        |             |  |
| RL        | Nat. Genet. 21:385-389(1999).  |           |      |     |     |            |        |             |  |
| RN        | [2]  |           |      |     |     |            |        |             |  |
| RP        | SEQUENCE FROM N.A.   |           |      |     |     |            |        |             |  |
| RC        | STRAIN=AR39;   |           |      |     |     |            |        |             |  |
| RX        | MEDLINE=20150255; PubMed=10684935;   |           |      |     |     |            |        |             |  |
| RA        | Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  |           |      |     |     |            |        |             |  |
| RA        | White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,   |           |      |     |     |            |        |             |  |
| RA        | Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,   |           |      |     |     |            |        |             |  |
| RA        | Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,   |           |      |     |     |            |        |             |  |
| RA        | Eisen J., Fraser C.M.;   |           |      |     |     |            |        |             |  |
| RT        | "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  |           |      |     |     |            |        |             |  |
| RL        | pneumoniae AR39.";   |           |      |     |     |            |        |             |  |
| RN        | [3]  |           |      |     |     |            |        |             |  |
| RP        | SEQUENCE FROM N.A.   |           |      |     |     |            |        |             |  |
| RC        | STRAIN=J138;   |           |      |     |     |            |        |             |  |
| RX        | MEDLINE=20330349; PubMed=10871362;   |           |      |     |     |            |        |             |  |
| RA        | Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,   |           |      |     |     |            |        |             |  |
| RA        | Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  |           |      |     |     |            |        |             |  |
| RT        | "Comparison of whole genome sequences of Chlamydia pneumoniae J138   |           |      |     |     |            |        |             |  |
| RL        | from Japan and CWL029 from USA.";  |           |      |     |     |            |        |             |  |
| RL        | Nucleic Acids Res. 28:2311-2314(2000).   |           |      |     |     |            |        |             |  |
| CC        | -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS  |           |      |     |     |            |        |             |  |
| CC        | INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR  |           |      |     |     |            |        |             |  |
| CC        | ACCURATE TRANSLATION (BY SIMILARITY).  |           |      |     |     |            |        |             |  |
| CC        | -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.  |           |      |     |     |            |        |             |  |
| CC        | This SWISS-PROT entry is copyright. It is produced through a collaboration   |           |      |     |     |            |        |             |  |
| CC        | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |           |      |     |     |            |        |             |  |
| CC        | the European Bioinformatics Institute. There are no restrictions on its  |           |      |     |     |            |        |             |  |
| CC        | use by non-profit institutions as long as its content is in no way   |           |      |     |     |            |        |             |  |
| CC        | modified and this statement is not removed. Usage by and for commercial  |           |      |     |     |            |        |             |  |
| CC        | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |           |      |     |     |            |        |             |  |
| CC        | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |           |      |     |     |            |        |             |  |
| CC        | -----  |           |      |     |     |            |        |             |  |
| DR        | EMBL; AE001593; AAD18233.1; -.   |           |      |     |     |            |        |             |  |
| DR        | EMBL; AE002228; AAF38503.1; -.   |           |      |     |     |            |        |             |  |

## ALIGNMENTS

```

DR EMBL; AP002545; BAA98290.1; -.
DR F1R; C72122; C72122.
DR F1R; H86500; H86500.
DR HSP; P02392; 1CTF.
DR PHCI-2DPAGE; Q929A1; -.
DR TIGR; CP0695; -.
DR HAMAP; MF 00368; -.
DR InterPro; IPR000206; Ribosomal_L12.
DR Pfam; PF00542; Ribosomal_L12; 1.
DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRfam; TIGR00855; L12; 1.
DR Ribosomal protein; Complete proteome.
FT INIT MET 0
FT CONFLICT 14 14 G -> R (IN REF. 1).
FT CONFLICT 51 51 A -> R (IN REF. 1).
FT CONFLICT 128 129 MISSING (IN REF. 1).
SQ SEQUENCE 128 AA; 13461 MW; 4E2F17A85B057CC CRC64;
Query Match 100.0%; Score 46; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
DB 1 TTESLETLVE 10

RESULT 2
RL7_CHLMU STANDARD; PRT; 129 AA.
ID RL7_CHLMU
AC P38001;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7 OR TC0590
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn;
RX MEDLINE=91008945; PubMed=2211507;
RA Engel J.N., Pollack J., Malik F., Ganem D.;
RT "Cloning and characterization of RNA polymerase core subunits of
Chlamydia trachomatis by using the polymerase chain reaction.";
RL J. Bacteriol. 172:5732-5741(1990).
[2]
RN 2
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nig9;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
ACCURATE TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002328; AAF39422.1; -.
DR F1R; H81684; H81684.
DR HSP; P02392; 1CTF.
DR Siena-2DPAGE; P38001; -.

```

```

DR TIGR; TC0590; -.
DR HAMAP; MF 00368; -.
DR InterPro; IPR000206; Ribosomal_L12.
DR Pfam; PF00542; Ribosomal_L12; 1.
DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRfam; TIGR00855; L12; 1.
DR Ribosomal protein; Complete proteome.
FT INIT MET 0
FT CONFLICT 14 14 G -> R (IN REF. 1).
FT CONFLICT 51 51 A -> R (IN REF. 1).
FT CONFLICT 128 129 MISSING (IN REF. 1).
SQ SEQUENCE 129 AA; 13441 MW; 27DBF2C6613DFB3 CRC64;
Query Match 100.0%; Score 46; DB 1; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
DB 1 TTESLETLVE 10

RESULT 3
RL7_CHLTR STANDARD; PRT; 129 AA.
ID RL7_CHLTR
AC O84318;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPL7 OR RL7 OR CT316.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
Mitchell W.P., Olinger L., Tatusov R.B., Zhao Q., Koonin E.V.,
Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.";
RT Chlamydia trachomatis.
RL Science 282:754-759(1998).
[2]
RN 2
RP SEQUENCE OF 1-10.
RC STRAIN=L2/434/Bu;
RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
Comanducci M., Christensen G., Birkelund S., Vretou E., Ratti G.,
Pallini V.;
RA Submitted (SEP-1994) to the SWISS-PROT data bank.
CC -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
ACCURATE TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001304; AAC67909.1; -.
DR F1R; A71530; A71530.
DR HSP; P02392; 1CTF.
DR PHCI-2DPAGE; O84318; -.
DR HAMAP; MF 00368; -.
DR InterPro; IPR000206; Ribosomal_L12.
DR Pfam; PF00542; Ribosomal_L12; 1.
DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRfam; TIGR00855; L12; 1.

```

KW Ribosomal protein; Complete proteome.

FT INIT MET 0  
SQ SEQUENCE 129 AA; 13439 MW; DFAFA383677FEB0 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 129;

Best Local Similarity 100.0%; Pred. No. 0.036; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0

QY 1 TTESLETLVE 10

Db 1 TTESLETLVE 10

#### RESULT 4

AMYG CANAL STANDARD; PRT; 946 AA.

ID ML34\_CANAL  
AC O74254;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
DE (1,4-alpha-D-glucan glucohydrolase).  
GN GAM1 OR GCA1.

OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxID=5476;

[1]\_TaxID=5476;

RN [1]\_TaxID=5476;

RP SEQUENCE FROM N.A.

RC STRAIN=SC5314;

RX MEDLINE=99451422; PubMed=10520161;

RA Sturtevant J., Dixon F., Wadsworth E., Latge J.-P., Zhao X.-J.,

RA "Identification and cloning of GCA1, a gene that encodes a cell

RT surface glucoamylase from *Candida albicans*."

RL Med. Mycol. 37:357-366 (1999).

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -!- SUBCELLULAR LOCATION: CELL WALL ASSOCIATED.

CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF082188; AAC31968.1; -

DR InterPro; IPR000322; Glyco\_hydro\_31.

DR Pfam; PF01055; Glyco\_hydro\_31; 1.

DR PROSITE; PS00129; GLYCOSYL HYDROL\_F31.1; 1.

DR PROSITE; PS00707; GLYCOSYL HYDROL\_F31.2; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;

Signal; Cell wall.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 946 GLUCOAMYLASE 1.

FT ACT\_SITE 462 462 BY SIMILARITY.

FT DOMAIN 519 532 SER/THR-RICH.

FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 570 570 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 895 895 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 912 912 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 946 AA; 105804 MW; BD6B640C4EEF1F70 CRC64;

Query Match 82.6%; Score 38; DB 1; Length 946;

Best Local Similarity 80.0%; Pred. No. 13;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

Db 318 TVESLETLVE 327

#### RESULT 5

ML34\_ARATH

ID ML34\_ARATH STANDARD; PRT; 316 AA.

AC Q9SSK7;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE MLP-like protein 34

GN MLP34 OR AtLG70850 OR F15H11.10.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

[1]\_TaxID=3702;

RN [1]\_TaxID=3702;

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Muller S., Klitt S., Hauser M.T.;

RT "Molecular and phylogenetic analysis of a gene family in Arabidopsis

RT thaliana with similarities to major latex, pathogenesis-related and

RT ripening-induced proteins."

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

RT thaliana."

RL Nature 408:816-820 (2000).

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the

RT SSP consortium (Salk/Stanford/PfGEC)."

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE MLP FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AJ306141; CAC83579.1; -  
 CC EMBL; AC008148; AAD55503.1; -  
 CC DR EMBL; AF372899; AAK49615.1; -  
 CC DR EMBL; AY057726; AAL15356.1; -  
 CC DR PIR; C96733; C96733.  
 CC DR Pfam; PF00407; Bet\_v\_1; 2.  
 CC KW Multigene family.  
 CC SQ SEQUENCE 316 AA; 35570 MW; E19EC47AE8AFC2A2 CRC64;

Query Match 78.3%; Score 36; DB 1; Length 316;  
 Best Local Similarity 80.0%; Pred. No. 9.5;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TTSETLIVE 10  
 |||||  
 Db 165 TTSETLIVE 174

## RESULT 6

TOLC VIBCH  
 ID TOLC VIBCH STANDARD; PRT; 438 AA.  
 AC Q9K2Y4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein tolc precursor.  
 GN TOLC OR VC2436.  
 GN Vibrio cholerae.  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RA Bina J.E., Mekalanos J.J.;  
 RT "Identification and characterization of Vibrio cholerae tolc";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae";  
 RL Nature 406:477-483 (2000).  
 CC -!- FUNCTION: MAY BE SPECIALIZED FOR SIGNAL SEQUENCE INDEPENDENT,  
 CC EXTRACELLULAR SECRETION IN GRAM-NEGATIVE BACTERIA (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PRF FAMILY OF SECRETION PROTEINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AF282892; AAF91468.1; -  
 CC EMBL; AE004313; AAF95579.1; -  
 CC DR PIR; B82077; B82077.  
 CC DR HSSP; P02930; 1EK9.  
 CC DR TIGR; VC2436; -  
 CC InterPro; IPR003423; OEP.

DR Pfam; PF02321; OEP; 2.  
 KW Transport; Outer membrane; Signal; Complete proteome.  
 FT SIGNAL 1..22 POTENTIAL.  
 FT CHAIN 23..438 OUTER MEMBRANE PROTEIN TOLC.  
 SQ SEQUENCE 438 AA; 47751 MW; 79BDDF309953CID5 CRC64;  
 Query Match 73.9%; Score 34; DB 1; Length 438;  
 Best Local Similarity 70.0%; Pred. No. 34;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 TTSETLIVE 10  
 |||||  
 Db 226 TTSETLIVE 235  
 RESULT 7  
 VB03\_VACCC  
 ID VB03\_VACCC STANDARD; PRT; 124 AA.  
 AC P21000;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein B3.  
 DE B3K.  
 GN Vaccinia virus (strain Copenhagen).  
 OS Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 OX NCBI\_TaxID=10249;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021027; PubMed=2219722;  
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RT "The complete DNA sequence of vaccinia virus";  
 RL Virology 179:247-266 (1990).  
 RN [2]  
 RP COMPLETE GENOME.  
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RT "Appendix to 'The complete DNA sequence of vaccinia virus'";  
 RL Virology 179:517-563 (1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M35027; AAA48199.1; -  
 CC DR PIR; B42526; B42526.  
 CC SQ SEQUENCE 124 AA; 14379 MW; 5CDBC949BC2FF692 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 124;  
 Best Local Similarity 77.8%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TESLETIVE 10  
 |||||  
 Db 94 TESLETIVE 102

## RESULT 8

RISB STAAM  
 ID RISB STAAM STANDARD; PRT; 154 AA.  
 AC Q931N8; Q99TAL1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)  
 DE (Lumazine synthase) (Riboflavin synthase beta chain).  
 GN RIBH OR SAV1767 OR SA1586 OR MW1708.

```

OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain WM2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamanoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (by similarity).
CC -!- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the DMRL synthase family.
CC -!- CAUTION: Ref.1 strain Mu50 sequence differs from that shown due to
CC a frameshift in position 28.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003363; BAB57929.1; ALT_FRAME.
DR EMBL; AP003363; BAB42854.1; -.
DR EMBL; AP004828; BAB95573.1; -.
DR PIR; A89962; A89962.
DR HAMAP; MF 00178; -.
DR InterPro; IPR002180; DMRL synthase.
DR Pfam; PF00885; DMRL synthase; 1.
DR ProDom; PD003664; DMRL synthase; 1.
DR TIGRFAMs; TIGR00114; ribH; 1.
KW Riboflavin biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 154 AA; 16410 MW; BC6AD39B6431BF44 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 154;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLTVE 10
DB 116 TTESIQAQVE 125

RESULT 9
DR EMBL; D11079; BAA01833.1; -.
DR PIR; JQ1797; JQ1797.

Staphylococcus aureus (strain Mu50 / ATCC 700699),
Staphylococcus aureus (strain N315), and
Staphylococcus aureus (strain WM2).
Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxID=158878, 158879, 196620;
SEQUENCE FROM N.A.
STRAIN=Mu50 / ATCC 700699, and N315;
MEDLINE=21311952; PubMed=11418146;
Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
"Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
Lancet 357:1225-1240(2001).
SEQUENCE FROM N.A.
STRAIN=WM2;
MEDLINE=22040717; PubMed=12044378;
Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
Yamanoto K., Hiramatsu K.;
"Genome and virulence determinants of high virulence community-
acquired MRSA.";
Lancet 359:1819-1827(2002).
FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-
phosphate yielding 6,7-dimethyl-8-lumazine (by similarity).
CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1-D-ribityl)lumazine =
riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
PATHWAY: Riboflavin biosynthesis; last step.
SIMILARITY: Belongs to the DMRL synthase family.
CAUTION: Ref.1 strain Mu50 sequence differs from that shown due to
a frameshift in position 28.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP003363; BAB57929.1; ALT_FRAME.
EMBL; AP003363; BAB42854.1; -.
EMBL; AP004828; BAB95573.1; -.
PIR; A89962; A89962.
HAMAP; MF 00178; -.
InterPro; IPR002180; DMRL synthase.
Pfam; PF00885; DMRL synthase; 1.
ProDom; PD003664; DMRL synthase; 1.
TIGRFAMs; TIGR00114; ribH; 1.
Riboflavin biosynthesis; Transferrase; Complete proteome.
SEQUENCE 154 AA; 16410 MW; BC6AD39B6431BF44 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 154;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLTVE 10
DB 116 TTESIQAQVE 125

RESULT 9
DR EMBL; D11079; BAA01833.1; -.
DR PIR; JQ1797; JQ1797.

```

```
SQ SEQUENCE 167 AA; 19410 MW; 82AP46891A7768D7 CRC64;
Query Match 71.7%; Score 33; DB 1; Length 167;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TESLETIVE 10
Db 94 TESLERLVE 102

RESULT 11
HTR2 HALVA STANDARD; PRT; 433 AA.
AC P42258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II) (Fragment).
GN HTRII.
OS Haloarcula vallismortis.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloarcula.
OX NCBI_TaxID=28442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 297115;
RX MEDLINE=9524074; PubMed=7708770;
RA Seidel R., Scharf B., Gautel M., Kleine K., Oesterheld D.,
RA Engelhard M.;
RT "The primary structure of sensory rhodopsin II: a member of an
RT additional retinal protein subgroup is coexpressed with its
RT transducer, the halobacterial transducer of rhodopsin II."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 1 HAMP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z35308; CAA84549.1; -
CC HSP; P02942; 1QV7.
CC InterPro; IPR004089; Chmtaxis_transd.
CC InterPro; IPR003660; HAMP.
CC InterPro; IPR004090; Me_Chemotaxis.
CC Pfam; PF00015; MCPsignal; 1.
CC PRINTS; PR00260; CHEMTRNSDUCK.
CC SMART; SM00304; HAMP; 1.
CC SMART; SM00283; HAMP; 1.
CC PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
CC PROSITE; PS00885; HAMP; 1.
KW Transducer; Photoreceptor; Transmembrane; Methylation.
FT NON_TER 1 1
FT DOMAIN 58 111 HAMP.
FT DOMAIN 130 366 METHYL-ACCEPTING_TRANSDUCER.
SQ SEQUENCE 433 AA; 45935 MW; 90507B897D943C0 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 433;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETIVE 10
Db 323 TVDALETIVE 332
```

```
RESULT 12
POLG DENA STANDARD; PRT; 3386 AA.
AC P09866; Q88661; Q88662; Q88663; Q88664; Q88665; Q88666; Q88667;
AC Q88668; Q88669; Q88670; Q88671;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope glycoprotein M); Major envelope protein E;
DE Nonstructural proteins NS1, NS2, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)].
DE Dengue virus type 4.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11070;
RN [1]
RP SEQUENCE OF 1-776 FROM N.A.
RX MEDLINE=8704106; PubMed=3022479;
RA Zhao B., Mackow E., Buckler-White A., Markoff L., Chancok R.M.,
RA Lai C.-J., Makino Y.;
RT "Cloning full-length dengue type 4 viral DNA sequences: analysis of
RT genes coding for structural proteins."
RL Virology 155:77-88(1986).
RN [2]
RP SEQUENCE OF 774-3386 FROM N.A.
RX MEDLINE=87293881; PubMed=3039728;
RA Mackow E., Makino Y., Zhao B., Zhang Y.M., Markoff L.,
RA Buckler-White A., Guiler M., Chanock R., Lai C.-J.;
RT "The nucleotide sequence of dengue type 4 virus: analysis of genes
RT coding for nonstructural proteins."
RL Virology 159:217-228(1987).
RN [3]
RP PROCESSING OF THE M PROTEIN.
RC STRAIN=814669;
RX MEDLINE=89311624; PubMed=2501515;
RA Markoff L.;
RT "In vitro processing of dengue virus structural proteins: cleavage of
RT the pre-membrane protein."
RL J. Virol. 63:3345-3352(1989).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14931; AAA42964.1; -
CC MEROPS; S07.002; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001122; Flavi_capsidC.
CC InterPro; IPR000336; Flavi_glycoprote.
CC InterPro; IPR001850; Flavi_helicase.
CC InterPro; IPR000069; Flavi_M.
CC InterPro; IPR001157; Flavi_NS1.
CC InterPro; IPR000752; Flavi_NS2A.
```



DR InterPro: IPR000487; Flavi\_NS2B.  
DR InterPro: IPR000404; Flavi\_NS4B.  
DR InterPro: IPR001528; Flavi\_NS4B.  
DR InterPro: IPR000208; Flavi\_NS5.  
DR InterPro: IPR002535; Flavi\_propep.  
DR InterPro: IPR002877; Ftsu.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01003; Flavi\_capsid; 1.  
DR Pfam: PF02832; Flavi\_glycop.C; 1.  
DR Pfam: PF00869; Flavi\_glycoprot; 1.  
DR Pfam: PF00949; Flavi\_helicase; 1.  
DR Pfam: PF01004; Flavi\_M; 1.  
DR Pfam: PF00948; Flavi\_NS1; 1.  
DR Pfam: PF01005; Flavi\_NS2A; 1.  
DR Pfam: PF01002; Flavi\_NS2B; 1.  
DR Pfam: PF01350; Flavi\_NS4A; 1.  
DR Pfam: PF01349; Flavi\_NS4B; 1.  
DR Pfam: PF00972; Flavi\_NS5; 1.  
DR Pfam: PF01570; Flavi\_propep; 1.  
DR Pfam: PF01728; Ftsu; 1.  
DR ProDom: PD001556; Flavi\_glycoprotE; 1.  
DR ProDom: PD001496; Flavi\_NS1; 1.  
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;  
KW ATP-binding; Transmembrane; Nonstructural protein.  
FT CHAIN 1 113 CAPSID PROTEIN C.  
FT PROPEP 114 204  
FT CHAIN 205 279  
FT CHAIN 280 773  
FT CHAIN 774 1184  
FT CHAIN 1185 1343  
FT CHAIN 1344 1473  
FT CHAIN 1474 2091  
FT CHAIN 2092 2374  
FT CHAIN 2375 2486  
FT CHAIN 2487 3386  
FT DOMAIN 377 390  
FT ACT\_SITE 1524 1524  
FT ACT\_SITE 1548 1548  
FT ACT\_SITE 1608 1608  
FT NP\_BIND 1666 1673  
FT SITE 1757 1760  
FT TRANSMEM 42 58  
FT TRANSMEM 267 283  
FT TRANSMEM 728 744  
FT TRANSMEM 753 769  
FT TRANSMEM 1157 1179  
FT DISULFID 282 309  
FT DISULFID 339 395  
FT DISULFID 353 384  
FT DISULFID 371 400  
FT DISULFID 464 564  
FT DISULFID 581 612  
FT CARBOHYD 182 182  
FT CARBOHYD 346 346  
FT CARBOHYD 432 432  
FT CARBOHYD 750 750  
FT CARBOHYD 903 903  
FT CARBOHYD 980 980  
FT CARBOHYD 2296 2296  
FT CARBOHYD 2300 2300  
FT CARBOHYD 2452 2452  
FT CARBOHYD 3386 3386  
SQ SEQUENCE 3386 AA; 378903 MW; 5A984B8742C54021 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 3386;

Best Local Similarity 87.5%; Pred. No. 5.1e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TESLETIV 9

|||||

2140 TESLETIM 2147

DB

RESULT 13  
ANK2 HUMAN  
ID ANK2\_HUMAN STANDARD; PRT; 3924 AA.  
AC Q01484; Q01485;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).  
GN ANK2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RC TISSUE=Brain stem; PubMed=1830053;  
RX MEDLINE=91302466; PubMed=1830053;  
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;  
RT "Isolation and characterization of cDNAs encoding human brain  
RT ankyrins reveal a family of alternatively spliced genes.";  
RL J. Cell Biol. 114:241-253(1991).  
RN [2]  
RP REVISIONS.  
RA Carpenter S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain stem; PubMed=8253844;  
RX MEDLINE=94075409; PubMed=8253844;  
RA Chan W., Kordeli E., Bennett V.;  
RT "440-kD ankyrinB: structure of the major developmentally regulated  
RT domain and selective localization in unmyelinated axons.";  
RL J. Cell Biol. 123:1463-1473(1993).  
RN [4]  
RP SEQUENCE OF 463-495 FROM N.A.  
RX MEDLINE=92009921; PubMed=1833308;  
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,  
RA Lux S.E., Ward D.C., Forget B.G.;  
RT "Isolation and chromosomal localization of a novel nonerythroid  
RT ankyrin gene.";  
RL Genomics 10:858-866(1991).  
CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal  
CC elements. Also bind to cytoskeletal proteins.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC Name=2;  
CC Name=3;  
CC IsoId=Q01484-1; Sequence=Displayed;  
CC IsoId=Q01484-2; Sequence=VSP\_000267, VSP\_000268;  
CC IsoId=Q01484-3; Sequence=VSP\_000268;  
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL  
CC CELLS THROUGHOUT THE BRAIN.  
CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES  
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE  
CC -!- SIMILARITY: Contains 23 ANK repeats.  
CC -!- SIMILARITY: Contains 1 death domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X56957; CAA40278.1; -.  
CC EMBL; X56958; CAA40279.2; -.  
CC EMBL; Z26634; CAA42644.1; -.  
CC EMBL; M37123; AAA62828.1; -.  
CC PIR; S37431; S37431.

DR HSPSP; P42771; 1DC2.  
 DR Genew; HGNC:493; ANK2.  
 DR MIM; 106410; -  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000488; Death.  
 DR InterPro; IPR000906; ZUS.  
 DR Pfam; PF00023; ank; 23  
 DR Pfam; PF00531; death; 1.  
 DR Pfam; PF00791; ZUS; 1.  
 DR PRINTS; PR01415; ANKYRIN.  
 DR SMART; SM00248; ANK; 22.  
 DR SMART; SM00005; DEATH; 1.  
 DR SMART; SM00218; ZUS; 1.  
 DR PROSITE; PS50088; ANK\_REPEAT; 20.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS50017; DEATH\_DOMAIN; 1.  
 KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;  
 KW Phosphorylation.  
 REPEAT 63 92 ANK 1.  
 REPEAT 96 125 ANK 2.  
 REPEAT 129 158 ANK 3.  
 REPEAT 162 191 ANK 4.  
 REPEAT 193 220 ANK 5.  
 REPEAT 232 261 ANK 6.  
 REPEAT 265 294 ANK 7.  
 REPEAT 298 327 ANK 8.  
 REPEAT 331 360 ANK 9.  
 REPEAT 364 393 ANK 10.  
 REPEAT 397 426 ANK 11.  
 REPEAT 430 459 ANK 12.  
 REPEAT 463 492 ANK 13.  
 REPEAT 496 525 ANK 14.  
 REPEAT 529 558 ANK 15.  
 REPEAT 562 591 ANK 16.  
 REPEAT 595 624 ANK 17.  
 REPEAT 628 657 ANK 18.  
 REPEAT 661 690 ANK 19.  
 REPEAT 694 723 ANK 20.  
 REPEAT 727 756 ANK 21.  
 REPEAT 760 789 ANK 22.  
 REPEAT 793 822 ANK 23.  
 REPEAT 825 854 ANK 24.  
 REPEAT 857 886 ANK 25.  
 REPEAT 889 918 ANK 26.  
 REPEAT 921 950 ANK 27.  
 REPEAT 953 982 ANK 28.  
 REPEAT 985 1014 ANK 29.  
 REPEAT 1017 1046 ANK 30.  
 REPEAT 1049 1078 ANK 31.  
 REPEAT 1081 1110 ANK 32.  
 REPEAT 1113 1142 ANK 33.  
 REPEAT 1145 1174 ANK 34.  
 REPEAT 1177 1206 ANK 35.  
 REPEAT 1209 1238 ANK 36.  
 REPEAT 1241 1270 ANK 37.  
 REPEAT 1273 1302 ANK 38.  
 REPEAT 1305 1334 ANK 39.  
 REPEAT 1337 1366 ANK 40.  
 REPEAT 1369 1398 ANK 41.  
 REPEAT 1401 1430 ANK 42.  
 REPEAT 1433 1462 ANK 43.  
 REPEAT 1465 1494 ANK 44.  
 REPEAT 1497 1526 ANK 45.  
 REPEAT 1529 1558 ANK 46.  
 REPEAT 1561 1590 ANK 47.  
 REPEAT 1593 1622 ANK 48.  
 REPEAT 1625 1654 ANK 49.  
 REPEAT 1657 1686 ANK 50.  
 REPEAT 1689 1718 ANK 51.  
 REPEAT 1721 1750 ANK 52.  
 REPEAT 1753 1782 ANK 53.  
 REPEAT 1785 1814 ANK 54.  
 REPEAT 1817 1846 ANK 55.  
 REPEAT 1849 1878 ANK 56.  
 REPEAT 1881 1910 ANK 57.  
 REPEAT 1913 1942 ANK 58.  
 REPEAT 1945 1974 ANK 59.  
 REPEAT 1977 2006 ANK 60.  
 REPEAT 2009 2038 ANK 61.  
 REPEAT 2041 2070 ANK 62.  
 REPEAT 2073 2102 ANK 63.  
 REPEAT 2105 2134 ANK 64.  
 REPEAT 2137 2166 ANK 65.  
 REPEAT 2169 2198 ANK 66.  
 REPEAT 2201 2230 ANK 67.  
 REPEAT 2233 2262 ANK 68.  
 REPEAT 2265 2294 ANK 69.  
 REPEAT 2297 2326 ANK 70.  
 REPEAT 2329 2358 ANK 71.  
 REPEAT 2361 2390 ANK 72.  
 REPEAT 2393 2422 ANK 73.  
 REPEAT 2425 2454 ANK 74.  
 REPEAT 2457 2486 ANK 75.  
 REPEAT 2489 2518 ANK 76.  
 REPEAT 2521 2550 ANK 77.  
 REPEAT 2553 2582 ANK 78.  
 REPEAT 2585 2614 ANK 79.  
 REPEAT 2617 2646 ANK 80.  
 REPEAT 2649 2678 ANK 81.  
 REPEAT 2681 2710 ANK 82.  
 REPEAT 2713 2742 ANK 83.  
 REPEAT 2745 2774 ANK 84.  
 REPEAT 2777 2806 ANK 85.  
 REPEAT 2809 2838 ANK 86.  
 REPEAT 2841 2870 ANK 87.  
 REPEAT 2873 2902 ANK 88.  
 REPEAT 2905 2934 ANK 89.  
 REPEAT 2937 2966 ANK 90.  
 REPEAT 2969 2998 ANK 91.  
 REPEAT 3001 3030 ANK 92.  
 REPEAT 3033 3062 ANK 93.  
 REPEAT 3065 3094 ANK 94.  
 REPEAT 3097 3126 ANK 95.  
 REPEAT 3129 3158 ANK 96.  
 REPEAT 3161 3190 ANK 97.  
 REPEAT 3193 3222 ANK 98.  
 REPEAT 3225 3254 ANK 99.  
 REPEAT 3257 3286 ANK 100.  
 REPEAT 3289 3318 ANK 101.  
 REPEAT 3321 3350 ANK 102.  
 REPEAT 3353 3382 ANK 103.  
 REPEAT 3385 3414 ANK 104.  
 REPEAT 3417 3446 ANK 105.  
 REPEAT 3449 3478 ANK 106.  
 REPEAT 3481 3510 ANK 107.  
 REPEAT 3513 3542 ANK 108.  
 REPEAT 3545 3574 ANK 109.  
 REPEAT 3577 3606 ANK 110.  
 REPEAT 3609 3638 ANK 111.  
 REPEAT 3641 3670 ANK 112.  
 REPEAT 3673 3702 ANK 113.  
 REPEAT 3705 3734 ANK 114.  
 REPEAT 3737 3766 ANK 115.  
 REPEAT 3769 3798 ANK 116.  
 REPEAT 3801 3830 ANK 117.  
 REPEAT 3833 3862 ANK 118.  
 REPEAT 3865 3894 ANK 119.  
 REPEAT 3897 3926 ANK 120.  
 REPEAT 3929 3958 ANK 121.  
 REPEAT 3961 3990 ANK 122.  
 REPEAT 3993 4022 ANK 123.  
 REPEAT 4025 4054 ANK 124.  
 REPEAT 4057 4086 ANK 125.  
 REPEAT 4089 4118 ANK 126.  
 REPEAT 4121 4150 ANK 127.  
 REPEAT 4153 4182 ANK 128.  
 REPEAT 4185 4214 ANK 129.  
 REPEAT 4217 4246 ANK 130.  
 REPEAT 4249 4278 ANK 131.  
 REPEAT 4281 4310 ANK 132.  
 REPEAT 4313 4342 ANK 133.  
 REPEAT 4345 4374 ANK 134.  
 REPEAT 4377 4406 ANK 135.  
 REPEAT 4409 4438 ANK 136.  
 REPEAT 4441 4470 ANK 137.  
 REPEAT 4473 4502 ANK 138.  
 REPEAT 4505 4534 ANK 139.  
 REPEAT 4537 4566 ANK 140.  
 REPEAT 4569 4598 ANK 141.  
 REPEAT 4601 4630 ANK 142.  
 REPEAT 4633 4662 ANK 143.  
 REPEAT 4665 4694 ANK 144.  
 REPEAT 4697 4726 ANK 145.  
 REPEAT 4729 4758 ANK 146.  
 REPEAT 4761 4790 ANK 147.  
 REPEAT 4793 4822 ANK 148.  
 REPEAT 4825 4854 ANK 149.  
 REPEAT 4857 4886 ANK 150.  
 REPEAT 4889 4918 ANK 151.  
 REPEAT 4921 4950 ANK 152.  
 REPEAT 4953 4982 ANK 153.  
 REPEAT 4985 5014 ANK 154.  
 REPEAT 5017 5046 ANK 155.  
 REPEAT 5049 5078 ANK 156.  
 REPEAT 5081 5110 ANK 157.  
 REPEAT 5113 5142 ANK 158.  
 REPEAT 5145 5174 ANK 159.  
 REPEAT 5177 5206 ANK 160.  
 REPEAT 5209 5238 ANK 161.  
 REPEAT 5241 5270 ANK 162.  
 REPEAT 5273 5302 ANK 163.  
 REPEAT 5305 5334 ANK 164.  
 REPEAT 5337 5366 ANK 165.  
 REPEAT 5369 5398 ANK 166.  
 REPEAT 5401 5430 ANK 167.  
 REPEAT 5433 5462 ANK 168.  
 REPEAT 5465 5494 ANK 169.  
 REPEAT 5497 5526 ANK 170.  
 REPEAT 5529 5558 ANK 171.  
 REPEAT 5561 5590 ANK 172.  
 REPEAT 5593 5622 ANK 173.  
 REPEAT 5625 5654 ANK 174.  
 REPEAT 5657 5686 ANK 175.  
 REPEAT 5689 5718 ANK 176.  
 REPEAT 5721 5750 ANK 177.  
 REPEAT 5753 5782 ANK 178.  
 REPEAT 5785 5814 ANK 179.  
 REPEAT 5817 5846 ANK 180.  
 REPEAT 5849 5878 ANK 181.  
 REPEAT 5881 5910 ANK 182.  
 REPEAT 5913 5942 ANK 183.  
 REPEAT 5945 5974 ANK 184.  
 REPEAT 5977 6006 ANK 185.  
 REPEAT 6009 6038 ANK 186.  
 REPEAT 6041 6070 ANK 187.  
 REPEAT 6073 6102 ANK 188.  
 REPEAT 6105 6134 ANK 189.  
 REPEAT 6137 6166 ANK 190.  
 REPEAT 6169 6198 ANK 191.  
 REPEAT 6201 6230 ANK 192.  
 REPEAT 6233 6262 ANK 193.  
 REPEAT 6265 6294 ANK 194.  
 REPEAT 6297 6326 ANK 195.  
 REPEAT 6329 6358 ANK 196.  
 REPEAT 6361 6390 ANK 197.  
 REPEAT 6393 6422 ANK 198.  
 REPEAT 6425 6454 ANK 199.  
 REPEAT 6457 6486 ANK 200.  
 REPEAT 6489 6518 ANK 201.  
 REPEAT 6521 6550 ANK 202.  
 REPEAT 6553 6582 ANK 203.  
 REPEAT 6585 6614 ANK 204.  
 REPEAT 6617 6646 ANK 205.  
 REPEAT 6649 6678 ANK 206.  
 REPEAT 6681 6710 ANK 207.  
 REPEAT 6713 6742 ANK 208.  
 REPEAT 6745 6774 ANK 209.  
 REPEAT 6777 6806 ANK 210.  
 REPEAT 6809 6838 ANK 211.  
 REPEAT 6841 6870 ANK 212.  
 REPEAT 6873 6902 ANK 213.  
 REPEAT 6905 6934 ANK 214.  
 REPEAT 6937 6966 ANK 215.  
 REPEAT 6969 6998 ANK 216.  
 REPEAT 7001 7030 ANK 217.  
 REPEAT 7033 7062 ANK 218.  
 REPEAT 7065 7094 ANK 219.  
 REPEAT 7097 7126 ANK 220.  
 REPEAT 7129 7158 ANK 221.  
 REPEAT 7161 7190 ANK 222.  
 REPEAT 7193 7222 ANK 223.  
 REPEAT 7225 7254 ANK 224.  
 REPEAT 7257 7286 ANK 225.  
 REPEAT 7289 7318 ANK 226.  
 REPEAT 7321 7350 ANK 227.  
 REPEAT 7353 7382 ANK 228.  
 REPEAT 7385 7414 ANK 229.  
 REPEAT 7417 7446 ANK 230.  
 REPEAT 7449 7478 ANK 231.  
 REPEAT 7481 7510 ANK 232.  
 REPEAT 7513 7542 ANK 233.  
 REPEAT 7545 7574 ANK 234.  
 REPEAT 7577 7606 ANK 235.  
 REPEAT 7609 7638 ANK 236.  
 REPEAT 7641 7670 ANK 237.  
 REPEAT 7673 7702 ANK 238.  
 REPEAT 7705 7734 ANK 239.  
 REPEAT 7737 7766 ANK 240.  
 REPEAT 7769 7798 ANK 241.  
 REPEAT 7801 7830 ANK 242.  
 REPEAT 7833 7862 ANK 243.  
 REPEAT 7865 7894 ANK 244.  
 REPEAT 7897 7926 ANK 245.  
 REPEAT 7929 7958 ANK 246.  
 REPEAT 7961 7990 ANK 247.  
 REPEAT 7993 8022 ANK 248.  
 REPEAT 8025 8054 ANK 249.  
 REPEAT 8057 8086 ANK 250.  
 REPEAT 8089 8118 ANK 251.  
 REPEAT 8121 8150 ANK 252.  
 REPEAT 8153 8182 ANK 253.  
 REPEAT 8185 8214 ANK 254.  
 REPEAT 8217 8246 ANK 255.  
 REPEAT 8249 8278 ANK 256.  
 REPEAT 8281 8310 ANK 257.  
 REPEAT 8313 8342 ANK 258.  
 REPEAT 8345 8374 ANK 259.  
 REPEAT 8377 8406 ANK 260.  
 REPEAT 8409 8438 ANK 261.  
 REPEAT 8441 8470 ANK 262.  
 REPEAT 8473 8502 ANK 263.  
 REPEAT 8505 8534 ANK 264.  
 REPEAT 8537 8566 ANK 265.  
 REPEAT 8569 8598 ANK 266.  
 REPEAT 8601 8630 ANK 267.  
 REPEAT 8633 8662 ANK 268.  
 REPEAT 8665 8694 ANK 269.  
 REPEAT 8697 8726 ANK 270.  
 REPEAT 8729 8758 ANK 271.  
 REPEAT 8761 8790 ANK 272.  
 REPEAT 8793 8822 ANK 273.  
 REPEAT 8825 8854 ANK 274.  
 REPEAT 8857 8886 ANK 275.  
 REPEAT 8889 8918 ANK 276.  
 REPEAT 8921 8950 ANK 277.  
 REPEAT 8953 8982 ANK 278.  
 REPEAT 8985 9014 ANK 279.  
 REPEAT 9017 9046 ANK 280.  
 REPEAT 9049 9078 ANK 281.  
 REPEAT 9081 9110 ANK 282.  
 REPEAT 9113 9142 ANK 283.  
 REPEAT 9145 9174 ANK 284.  
 REPEAT 9177 9206 ANK 285.  
 REPEAT 9209 9238 ANK 286.  
 REPEAT 9241 9270 ANK 287.  
 REPEAT 9273 9302 ANK 288.  
 REPEAT 9305 9334 ANK 289.  
 REPEAT 9337 9366 ANK 290.  
 REPEAT 9369 9398 ANK 291.  
 REPEAT 9401 9430 ANK 292.  
 REPEAT 9433 9462 ANK 293.  
 REPEAT 9465 9494 ANK 294.  
 REPEAT 9497 9526 ANK 295.  
 REPEAT 9529 9558 ANK 296.  
 REPEAT 9561 9590 ANK 297.  
 REPEAT 9593 9622 ANK 298.  
 REPEAT 9625 9654 ANK 299.  
 REPEAT 9657 9686 ANK 300.  
 REPEAT 9689 9718 ANK 301.  
 REPEAT 9721 9750 ANK 302.  
 REPEAT 9753 9782 ANK 303.  
 REPEAT 9785 9814 ANK 304.  
 REPEAT 9817 9846 ANK 305.  
 REPEAT 9849 9878 ANK 306.  
 REPEAT 9881 9910 ANK 307.  
 REPEAT 9913 9942 ANK 308.  
 REPEAT 9945 9974 ANK 309.  
 REPEAT 9977 10006 ANK 310.  
 REPEAT 10009 10038 ANK 311.  
 REPEAT 10041 10070 ANK 312.  
 REPEAT 10073 10102 ANK 313.  
 REPEAT 10105 10134 ANK 314.  
 REPEAT 10137 10166 ANK 315.  
 REPEAT 10169 10198 ANK 316.  
 REPEAT 10201 10230 ANK 317.  
 REPEAT 10233 10262 ANK 318.  
 REPEAT 10265 10294 ANK 319.  
 REPEAT 10297 10326 ANK 320.  
 REPEAT 10329 10358 ANK 321.  
 REPEAT 10361 10390 ANK 322.  
 REPEAT 10393 10422 ANK 323.  
 REPEAT 10425 10454 ANK 324.  
 REPEAT 10457 10486 ANK 325.  
 REPEAT 10489 10518 ANK 326.  
 REPEAT 10521 10550 ANK 327.  
 REPEAT 10553 10582 ANK 328.  
 REPEAT 10585 10614 ANK 329.  
 REPEAT 10617 10646 ANK 330.  
 REPEAT 10649 10678 ANK 331.  
 REPEAT 10681 10710 ANK 332.  
 REPEAT 10713 10742 ANK 333.  
 REPEAT 10745 10774 ANK 334.  
 REPEAT 10777 10806 ANK 335.  
 REPEAT 10809 10838 ANK 336.  
 REPEAT 10841 10870 ANK 337.  
 REPEAT 10873 10902 ANK 338.  
 REPEAT 10905 10934 ANK 339.  
 REPEAT 10937 10966 ANK 340.  
 REPEAT 10969 10998 ANK 341.  
 REPEAT 11001 11030 ANK 342.  
 REPEAT 11033 11062 ANK 343.  
 REPEAT 11065 11094 ANK 344.  
 REPEAT 11097 11126 ANK 345.  
 REPEAT 11129 11158 ANK 346.  
 REPEAT 11161 11190 ANK 347.  
 REPEAT 11193 11222 ANK 348.  
 REPEAT 11225 11254 ANK 349.  
 REPEAT 11257 11286 ANK 350.  
 REPEAT 11289 11318 ANK 351.  
 REPEAT 11321 11350 ANK 352.  
 REPEAT 11353 11382 ANK 353.  
 REPEAT 11385 11414 ANK 354.  
 REPEAT 11417 11446 ANK 355.  
 REPEAT 11449 11478 ANK 356.  
 REPEAT 11481 11510 ANK 357.  
 REPEAT 11513 11542 ANK 358.  
 REPEAT 11545 11574 ANK 359.  
 REPEAT 11577 11606 ANK 360.  
 REPEAT 11609 11638 ANK 361.  
 REPEAT 11641 11670 ANK 362.  
 REPEAT 11673 11702 ANK 363.  
 REPEAT 11705 11734 ANK 364.  
 REPEAT 11737 11766 ANK 365.  
 REPEAT 11769 11798 ANK 366.  
 REPEAT 11801 11830 ANK 367.  
 REPEAT 11833 11862 ANK 368.  
 REPEAT 11865 11894 ANK 369.  
 REPEAT 11897 11926 ANK 370.  
 REPEAT 11929 11958 ANK 371.  
 REPEAT 11961 11990 ANK 372.  
 REPEAT 11993 12022 ANK 373.  
 REPEAT 12025 12054 ANK 374.  
 REPEAT 12057 12086 ANK 375.  
 REPEAT 12089 12118 ANK 376.  
 REPEAT 12121 12150 ANK 377.  
 REPEAT 12153 12182 ANK 378.  
 REPEAT 12185 12214 ANK 379.  
 REPEAT 12217 12246 ANK 380.  
 REPEAT 12249 12278 ANK 381.  
 REPEAT 12281 12310 ANK 382.  
 REPEAT 12313 12342 ANK 383.  
 REPEAT 12345 12374 ANK 384.  
 REPEAT 12377 12406 ANK 385.  
 REPEAT 12409 12438 ANK 386.  
 REPEAT 12441 12470 ANK 387.  
 REPEAT 12473 12502 ANK 388.  
 REPEAT 12505 12534 ANK 389.  
 REPEAT 12537 12566 ANK 390.  
 REPEAT 12569 12598 ANK 391.  
 REPEAT 12601 12630 ANK 392.  
 REPEAT 12633 12662 ANK 393.  
 REPEAT 12665 12694 ANK 394.  
 REPEAT 12697 12726 ANK 395.  
 REPEAT 12729 12758 ANK 396.  
 REPEAT 12761 12790 ANK 397.  
 REPEAT 12793 12822 ANK 398.  
 REPEAT 12825 12854 ANK 399.  
 REPEAT 12857 12886 ANK 400.  
 REPEAT 12889 12918 ANK 401.  
 REPEAT 12921 12950 ANK 402.  
 REPEAT 12953 12982 ANK 403.  
 REPEAT 12985 13014 ANK 404.  
 REPEAT 13017 13046 ANK 405.  
 REPEAT 13049 13078 ANK 406.  
 REPEAT 13081 13110 ANK 407.  
 REPEAT 13113 13142 ANK 408.  
 REPEAT 13145 13174 ANK 409.  
 REPEAT 13177 13206 ANK 410.  
 REPEAT 13209 13238 ANK 411.  
 REPEAT 13241 13270 ANK 412.  
 REPEAT 13273 13302 ANK 413.  
 REPEAT 13305 13334 ANK 414.  
 REPEAT 13337 13366 ANK 415.  
 REPEAT 13369 13398 ANK 416.  
 REPEAT 13401 13430 ANK 417.  
 REPEAT 13433 13462 ANK 418.  
 REPEAT 13465 13494 ANK 419.  
 REPEAT 13497 13526 ANK 420.  
 REPEAT 13529 13558 ANK 421.  
 REPEAT 13561 13590 ANK 422.  
 REPEAT 13593 13622 ANK 423.  
 REPEAT 13625 13654 ANK 424.  
 REPEAT 13657 13686 ANK 425.  
 REPEAT 13689 13718 ANK 426.  
 REPEAT 13721 13750 ANK 427.  
 REPEAT 13753 13782 ANK 428.  
 REPEAT 13785 13814 ANK 429.  
 REPEAT 13817 13846 ANK 430.  
 REPEAT 13849 13878 ANK 431.  
 REPEAT 13881 13910 ANK 432.  
 REPEAT 13913 13942 ANK 433.  
 REPEAT 13945 13974 ANK 434.  
 REPEAT 13977 14006 ANK 435.  
 REPEAT 14009 14038 ANK 436.  
 REPEAT 14041 14070 ANK 437.  
 REPEAT 14073 14102 ANK 438.  
 REPEAT 14105 14134 ANK 439.  
 REPEAT 14137 14166 ANK 440.  
 REPEAT 14169 14198 ANK 441.  
 REPEAT 14201 14230 ANK 442.  
 REPEAT 14233 14262 ANK 443.  
 REPEAT 14265 14294 ANK 444.  
 REPEAT 14297 14326 ANK 445.  
 REPEAT 14329 14358 ANK 446.  
 REPEAT 14361 14390 ANK 447.  
 REPEAT 14393 14422 ANK 448.  
 REPEAT 14425 14454 ANK 449.  
 REPEAT 14457 14486 ANK 450.  
 REPEAT 14489 14518 ANK 451.  
 REPEAT 14521 14550 ANK 452.  
 REPEAT 14553 14582 ANK 453.  
 REPEAT 14585 14614 ANK 454.  
 REPEAT 14617 14646 ANK 455.  
 REPEAT 14649 14678 ANK 456.  
 REPEAT 14681 14710 ANK 457.  
 REPEAT 14713 14742 ANK 458.  
 REPEAT 14745 14774 ANK 459.  
 REPEAT 14781 14810 ANK 460.  
 REPEAT 14813 14842 ANK 461.  
 REPEAT 14845 14874 ANK 462.  
 REPEAT 14877 14906 ANK 463.  
 REPEAT 14909 14938 ANK 464.  
 REPEAT 14941 14970 ANK 465.  
 REPEAT 14973 15002 ANK 466.  
 REPEAT 15005 15034 ANK 467.  
 REPEAT 15037 15066 ANK 468.  
 REPEAT 15069 15100 ANK 469.  
 REPEAT 15101 15130 ANK 470.  
 REPEAT 15133 15162 ANK 471.  
 REPEAT 15165 15194 ANK 472.  
 REPEAT 15197 15226 ANK 473.  
 REPEAT 15229 15258 ANK 474.  
 REPEAT 15261 15290 ANK 475.  
 REPEAT 15293 15322 ANK 476.  
 REPEAT 15325 15354 ANK 477.  
 REPEAT 15357 15386 ANK 478.  
 REPEAT 15389 15418 ANK 479.  
 REPEAT 15421 15450 ANK 480.  
 REPEAT 15453 15482 ANK 481.  
 REPEAT 15485 15514 ANK 482.  
 REPEAT 15517 15546 ANK 483.  
 REPEAT 15549 15578 ANK 484.  
 REPEAT 15581 15610 ANK 485.  
 REPEAT 15613 15642 ANK 486.  
 REPEAT 15645 15674 ANK 487.  
 REPEAT 15677 15706 ANK 488.  
 REPEAT 15709 15738 ANK 489.  
 REPEAT 15741 15770 ANK 490.  
 REPEAT 15773 15802 ANK 491.  
 REPEAT 15805 15834 ANK 492.  
 REPEAT 15837 15866 ANK 493.  
 REPEAT 15869 15898 ANK 494.  
 REPEAT 15901 15930 ANK 495.  
 REPEAT 15933 15962 ANK 496.  
 REPEAT 15965 15994 ANK 497.  
 REPEAT 15997 16026 ANK 498.  
 REPEAT 16029 16058 ANK 499.  
 REPEAT 16061 16090 ANK 500.  
 REPEAT 16093 16122 ANK 501.  
 REPEAT 16125 16154 ANK 502.  
 REPEAT 16157 16186 ANK 503.  
 REPEAT 16189 16218 ANK 504.  
 REPEAT 16221 16250 ANK 505.  
 REPEAT 16253 16282 ANK 506.  
 REPEAT 1628

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC CC MEMBRANE OF INFECTED CELLS.
CC CC -!- SUBCELLULAR LOCATION: Membrane-bound.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; K03347; AAA45372.1; -.
CC CC HSSP; P19554; 1VPU.
CC CC HIV; K03347; VPUS23.
CC CC InterPro; IPR002094; Vpu.
CC CC Pfam; PF00558; Vpu; 1.
CC CC Transmembrane; AIDS. 1
CC CC NON_TER 1
CC CC SEQUENCE 45 AA; 5342 MW; E9FF1EEAA174FA49 CRC64;
CC CC -----
CC CC Query Match 69.6%; Score 32; DB 1; Length 45;
CC CC Best Local Similarity 77.8%; Pred No. 6.8;
CC CC Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC CC -----
CC CC QY 2 TESLETVE 10
CC CC ||| |||||
CC CC Db 25 TEELSTLVE 33
CC CC -----
CC CC RESULT 16
CC CC RISB_ECOLI STANDARD; PRT; 156 AA.
CC CC ID RISB_ECOLI STANDARD; PRT; 156 AA.
CC CC AC P25540; P77114;
CC CC DT 01-MAY-1992 (Rel. 22, Created)
CC CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
CC CC DE (Lumazine synthase) (Riboflavin synthase beta chain).
CC CC GN RIBH OR RIBE OR B0415 OR C0525 OR Z0516 OR ECS0468 OR SF0352.
CC CC OS Escherichia coli,
CC CC OS Escherichia coli O6,
CC CC OS Escherichia coli O157:H7, and
CC CC OS Shigella flexneri.
CC CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC CC OC Enterobacteriaceae; Escherichia.
CC CC OX NCBI_TaxID=562, 217992, 83334, 623;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC SPECIES=E.coli; STRAIN=K12;
CC CC RX MEDLINE=93024316; PubMed=1406588;
CC CC RA Taura T., Ueguchi C., Shiba K., Ito K.;
CC CC RT "Insertional disruption of the nusB (ssyB) gene leads to
CC CC RT cold-sensitive growth of Escherichia coli and suppression of the
CC CC RT secY24 mutation.";
CC CC RL Mol. Gen. Genet. 234:429-432(1992).
CC CC RN [2]
CC CC RP SEQUENCE FROM N.A.
CC CC RC SPECIES=E.coli; STRAIN=K12 / MG1655;
CC CC RX MEDLINE=97426617; PubMed=9278503;
CC CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC CC RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC CC RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
CC CC RA Mau B., Shao Y.;
CC CC RT "The complete genome sequence of Escherichia coli K-12.";
CC CC RL Science 277:1453-1474(1997).
CC CC RN [3]
CC CC RP SEQUENCE FROM N.A.
CC CC RC SPECIES=E.coli;
CC CC RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
CC CC RA Duncan M., Federpspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
CC CC RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
CC CC RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC CC [4]

```

```

CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U11039; AAA95144.1; -.
CC CC EMBL; Z99112; CAB13590.1; -.
CC CC EMBL; Z99113; CAB13601.1; -.
CC CC FIR; A69679; A69679.
CC CC HSSP; P14687; 1AMU.
CC CC Subtilist; BG10930; pksK.
CC CC InterPro; IPR000873; AMP-bind.
CC CC InterPro; IPR001242; Condensatn.
CC CC InterPro; IPR000794; Ketoacyl-synt.
CC CC InterPro; IPR006163; Pp bind.
CC CC InterPro; IPR006162; Ppatne attach.
CC CC Pfam; PF00501; AMP-binding; 1
CC CC Pfam; PF00668; Condensation; 1.
CC CC Pfam; PF00109; ketoacyl-synt; 3.
CC CC Pfam; PF02801; ketoacyl-synt C; 3.
CC CC Pfam; PF00550; pp-binding; 5.
CC CC PRINTS; PR00154; AMPBINDING.
CC CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC CC PROSITE; PS00455; AMP_BINDING; 1.
CC CC PROSITE; PS00606; B KETOACYL SYNTHASE; 2.
CC CC PROSITE; PS00075; ACP DOMAIN; 5.
CC CC Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
CC CC Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
CC CC Complete proteome.
CC CC FT DOMAIN 1 68 ACYL CARRIER (ACP) 1.
CC CC FT DOMAIN 1063 1130 ACYL CARRIER (ACP) 2.
CC CC FT DOMAIN 2516 2589 ACYL CARRIER (ACP) 3.
CC CC FT DOMAIN 2618 2687 ACYL CARRIER (ACP) 4.
CC CC FT DOMAIN 3868 3937 ACYL CARRIER (ACP) 5.
CC CC FT BINDING 331 331 PHOSPHOPANTETHEINE (POTENTIAL).
CC CC FT BINDING 1093 1093 PHOSPHOPANTETHEINE (POTENTIAL).
CC CC FT BINDING 2552 2552 PHOSPHOPANTETHEINE (POTENTIAL).
CC CC FT BINDING 2650 2650 PHOSPHOPANTETHEINE (POTENTIAL).
CC CC FT ACT_SITE 2915 2915 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC CC FT BINDING 3900 3900 PHOSPHOPANTETHEINE (POTENTIAL).
CC CC FT ACT_SITE 4147 4147 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC CC SEQUENCE 4447 AA; 496058 MW; 9C6DB46AC0A9C057 CRC64;
CC CC -----
CC CC Query Match 71.7%; Score 33; DB 1; Length 4447;
CC CC Best Local Similarity 100.0%; Pred. No. 6.9e+02;
CC CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 TTESLET 7
CC CC |||||
CC CC Db 1294 TTESLET 1300
CC CC -----
CC CC RESULT 15
CC CC VPU_HV123 STANDARD; PRT; 45 AA.
CC CC AC P08805;
CC CC DT 01-NOV-1988 (Rel. 09, Created)
CC CC DT 01-NOV-1988 (Rel. 09, Last sequence update)
CC CC DE VPU protein (U ORF protein) (Fragment).
CC CC GN VPU.
CC CC OS Human immunodeficiency virus type 1 (Zaire 3 isolate) (HIV-1).
CC CC OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
CC CC OX NCBI_TaxID=11680;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RX MEDLINE=86259728; PubMed=3014529;
CC CC RA Willey R.W., Rutledge R.A., Dias S., Folks T., Theodore T.,
CC CC RA Buckler C.E., Martin M.A.;
CC CC RT "Identification of conserved and divergent domains within the
CC CC RT envelope gene of the acquired immunodeficiency syndrome retrovirus.";
CC CC RL Proc. Natl. Acad. Sci. U.S.A. 83:5038-5042(1986).
CC CC -!- FUNCTION: ACTS IN THE DEGRADATION OF CD4 IN THE ENDOPLASMIC
CC CC RETICULUM AND IN THE ENHANCEMENT OF VIRION RELEASE FROM THE PLASMA

```

RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=06:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postfal E.J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=0157:H7 / RMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 RT through comparison with genomes of *Escherichia coli* K12 and O157";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [8]  
 RP CHARACTERIZATION.  
 RC SPECIES=E.coli;  
 RX MEDLINE=97125954; PubMed=8969176;  
 RA Moertl S., Fischer M., Richter G., Tack J., Weinkauff S., Bacher A.;  
 RT "Biosynthesis of riboflavin. Lumazine synthase of *Escherichia coli*.";  
 RL J. Biol. Chem. 271:33201-33207(1996).  
 CC -|- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex  
 CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-  
 CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-  
 CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit  
 CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-  
 CC 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-  
 CC phosphate yielding 6,7-dimethyl-8-lumazine.  
 CC -|- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1'-D-ribityl)lumazine =  
 CC riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.  
 CC -|- PATHWAY: Riboflavin biosynthesis; last step.  
 CC -|- SUBUNIT: Oligomer of 60 beta subunits forming an icosahedral  
 CC capsid. The core of the capsid does not contain alpha subunits.  
 CC -|- SIMILARITY: Belongs to the DMRL synthase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X64395; CAA45736.1; -  
 CC EMBL; A5000148; AAC73518.1; -  
 CC EMBL; U82664; AAB40171.1; -  
 CC EMBL; A8016756; AAN79003.1; ALT INIT.  
 CC EMBL; A8005221; AAG54764.1; -  
 CC EMBL; AF002551; BAB33891.1; -  
 CC EMBL; A5015068; AAN42010.1; ALT\_INIT.  
 CC PIR; D90687; D90687.  
 CC PIR; H85537; H85537.  
 CC PIR; S26202; S26202.  
 CC HSSP; P11998; LRVV.  
 CC SWISS-2DPAGE; P25540; COLI.  
 CC EcoGene; EGI1322; ribH.  
 CC HAMAP; MF\_00178; -; 1.  
 CC InterPro; IPR002180; DMRL\_synthase.  
 CC Pfam; PF00885; DMRL\_synthase; 1.  
 CC ProDom; PD003664; DMRL\_synthase; 1.  
 CC TIGRFAMs; TIGR00114; ribH; 1.  
 CC Riboflavin biosynthesis; Transferase; Complete proteome.  
 KW SEQUENCE 156 AA; 16156 MW; 1F8504B2892195C7 CRC64;  
 SQ  
 Query Match 69.6%; Score 32; DB 1; Length 156;  
 Best Local Similarity 60.0%; Pred. No. 27;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TTSELTIVE 10  
 |||||:|  
 Db 118 TTSEIQE 127  
 RESULT 17  
 UCPI\_MOUSE  
 ID UCPI\_MOUSE STANDARD; PRT; 306 AA.  
 AC P12242;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Mitochondrial brown fat uncoupling protein 1 (UCP 1) (Thermogenin).  
 GN UCPI OR UCP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88315014; PubMed=3410843;  
 RA Kozak L.P., Britton J.H., Kozak U.C., Wells J.M.;  
 RT "The mitochondrial uncoupling protein gene. Correlation of exon  
 RT structure to transmembrane domains.";  
 RL J. Biol. Chem. 263:12274-12277(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SvJ;  
 RX MEDLINE=94088559; PubMed=8264627;  
 RA Kozak U.C., Kopecky J., Teisinger J., Enerback S., Boyer B.,  
 RA Kozak L.P.;  
 RT "An upstream enhancer regulating brown-fat-specific expression of the  
 RT mitochondrial uncoupling protein gene";  
 RL Mol. Cell. Biol. 14:59-67(1994).  
 CC -|- FUNCTION: UCP ARE MITOCHONDRIAL TRANSPORTER PROTEINS THAT CREATE  
 CC PROTON LEAKS ACROSS THE INNER MITOCHONDRIAL MEMBRANE, THUS  
 CC UNCOUPLING OXIDATIVE PHOSPHORYLATION FROM ATP SYNTHESIS. AS A  
 CC RESULT, ENERGY IS DISSIPATED IN THE FORM OF HEAT.  
 CC -|- SUBUNIT: ACTS AS A DIMER FORMING A PROTON CHANNEL.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.  
 CC -|- TISSUE SPECIFICITY: BROWN ADIPOSE TISSUE.  
 CC -|- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
 CC -|- SIMILARITY: Belongs to the mitochondrial carrier family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M21247; AAA40521.1; -;  
 DR EMBL; M21222; AAA40521.1; JOINED.  
 DR EMBL; M21244; AAA40521.1; JOINED.  
 DR EMBL; M21245; AAA40521.1; JOINED.  
 DR EMBL; M21246; AAA40521.1; JOINED.  
 DR EMBL; U63419; AAB05870.1; -;  
 DR EMBL; U63418; AAB07367.1; -;  
 DR PIR; A31106; A31106.  
 DR MGD; MGI:98894; UCP1.  
 DR GO; CO:0005739; C.mitochondrion; IDA.  
 DR InterPro; IPR002030; Mit uncoupling.  
 DR InterPro; IPR001993; Mitoch\_carrier.  
 DR Pfam; PF00153; mito\_carr; 3.  
 DR PRINTS; PR00784; MTUNCOUPLING.  
 DR PROSITE; PS00215; MITOCH\_CARRIER; 3.  
 DR Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT TRANSMEM 10 31 POTENTIAL.  
 FT TRANSMEM 73 95 POTENTIAL.  
 FT TRANSMEM 116 132 POTENTIAL.  
 FT TRANSMEM 178 194 POTENTIAL.  
 FT TRANSMEM 212 231 POTENTIAL.  
 FT TRANSMEM 266 288 POTENTIAL.  
 FT DOMAIN 273 295 PURINE NUCLEOTIDE BINDING  
 (BY SIMILARITY).  
 SQ SEQUENCE 306 AA; 33116 MW; C3FAC3D6B68F434F CRC64;

Query Match 69.6%; Score 32; DB 1; Length 306;  
 Best Local Similarity 87.5%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
 DB 165 TTESLSTL 172

RESULT 18  
 UCPI\_RAT STANDARD; PRT; 306 AA.  
 AC F04633;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mitochondrial brown fat uncoupling protein 1 (UCP 1) (Thermogenin).  
 GN UCPI OR UCP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1] POTENTIAL;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86111804; PubMed=3753702;  
 RA Bouillaud F., Weisenbach J., Ricquier D.;  
 RT "Complete cDNA-derived amino acid sequence of rat brown fat  
 RT uncoupling protein.";  
 RL J. Biol. Chem. 261:1487-1490 (1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86232450; PubMed=3012461;  
 RA Ridley R.G., Patel H.V., Gerber G.E., Morton R.C., Freeman K.B.;  
 RT "Complete nucleotide and derived amino acid sequence of cDNA encoding  
 RT the mitochondrial uncoupling protein of rat brown adipose tissue:  
 RT lack of a mitochondrial targeting presequence.";  
 RL Nucleic Acids Res. 14:4025-4035 (1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar;

RX MEDLINE=89076317; PubMed=3202878;  
 RA Bouillaud F., Raimbault S., Ricquier D.;  
 RT "The gene for rat uncoupling protein: complete sequence, structure of  
 RT primary transcript and evolutionary relationship between exons.";  
 RL Biochem. Biophys. Res. Commun. 157:783-792 (1988).  
 RN [4]  
 RP TOPOLOGY.  
 RX MEDLINE=94008980; PubMed=7691596;  
 RA Miroux B., Frossard V., Raimbault S., Ricquier D., Bouillaud F.;  
 RT "The topology of the brown adipose tissue mitochondrial uncoupling  
 RT protein determined with antibodies against its antigenic sites  
 RT revealed by a library of fusion proteins.";  
 RL EMBO J. 12:3739-3745 (1993).  
 CC -!- FUNCTION: UCP ARE MITOCHONDRIAL TRANSPORTER PROTEINS THAT CREATE  
 CC PROTON LEAKS ACROSS THE INNER MITOCHONDRIAL MEMBRANE, THUS  
 CC UNCOUPLING OXIDATIVE PHOSPHORYLATION FROM ATP SYNTHESIS. AS A  
 CC RESULT, ENERGY IS DISSIPATED IN THE FORM OF HEAT.  
 CC -!- SUBUNIT: ACTS AS A DIMER FORMING A PROTON CHANNEL.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.  
 CC -!- TISSUE SPECIFICITY: BROWN ADIPOSE TISSUE.  
 CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
 CC -!- SIMILARITY: Belongs to the mitochondrial carrier family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M11814; AAA19671.1; -;  
 DR EMBL; X03894; CAA27531.1; -;  
 DR EMBL; X12925; CAA31392.1; -;  
 DR PIR; A26294; A26294.  
 DR InterPro; IPR002030; Mit uncoupling.  
 DR InterPro; IPR001993; Mitoch\_carrier.  
 DR Pfam; PF00153; mito\_carr; 3.  
 DR PRINTS; PR00784; MTUNCOUPLING.  
 DR PROSITE; PS00215; MITOCH\_CARRIER; 3.  
 DR Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT TRANSMEM 10 31 POTENTIAL.  
 FT TRANSMEM 73 95 POTENTIAL.  
 FT TRANSMEM 116 132 POTENTIAL.  
 FT TRANSMEM 178 194 POTENTIAL.  
 FT TRANSMEM 212 231 POTENTIAL.  
 FT TRANSMEM 266 288 POTENTIAL.  
 FT DOMAIN 273 295 PURINE NUCLEOTIDE BINDING  
 (BY SIMILARITY).  
 SQ SEQUENCE 306 AA; 33080 MW; FB420AC67D2267A3 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 306;  
 Best Local Similarity 87.5%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
 DB 165 TTESLSTL 172

RESULT 19  
 ISDF\_HELPY STANDARD; PRT; 406 AA.  
 ID IDSF\_HELPY  
 AC O25664;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Isp/ispf bifunctional enzyme [includes: 2-C-methyl-D-erythritol 4-  
 DE phosphate cytidyltransferase (EC 2.7.7.60) (4-diphosphocytidyl-2-C-  
 DE methyl-D-erythritol synthase) (MEP cytidyltransferase) (MCT); 2-C-  
 DE methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)]

```

DE (MECPs) (MECDP-synthase)].
GN ISPDP OR HPI020.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]_STANDARD;
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kevlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq A.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: Bifunctional enzyme that catalyzes the formation of 4-
CC diphosphocytidyl-2C-methyl-D-erythritol from CTP and 2C-methyl-D-
CC erythritol 4-phosphate (ispD), and converts 4-diphosphocytidyl-2C-
CC methyl-D-erythritol 2-phosphate into 2C-methyl-D-erythritol 2,4-
CC cyclodiphosphate and CMP (ispF). Also converts 4-diphosphocytidyl-
CC 2C-methyl-D-erythritol into 2C-methyl-D-erythritol 3,4-
CC cyclodiphosphate and CMP (ispF) (By similarity).
CC -!- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
CC -!- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
CC + CMP.
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ISPDP FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ISPDP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000610; AAD08064.1; -.
CC TIGR; HPI020; -.
CC PIR; D64647; D64647.
CC DR HAMAP; MF_00107; fused; 1.
CC DR HAMAP; MF_00108; fused; 1.
CC DR InterPro; IPR001228; ISPDP synthase.
CC DR InterPro; IPR003526; YgBB.
CC DR Pfam; PF01128; ispD; 1.
CC DR Pfam; PF02542; YgBB; 1.
CC DR PIRSF; PIRSF006813; ispD/ispF synth; 1.
CC DR TIGRFAMs; TIGR00151; ispF; 1.
CC DR PROSITE; PS01295; ISPDP; 1.
CC DR PROSITE; PS01350; ISPDP; 1.
CC DR Transferase; Nucleotidyltransferase; Lyase; Isoprene biosynthesis;
CC Multifunctional enzyme; Complete proteome.
CC DOMAIN 1 244 406 CYCLODIPIHOSPHATE SYNTHASE.
CC FT FT 245 406 2-C-METHYL-D-ERYTHRITOL 2,4-
CC FT FT 244 406 CYCLODIPIHOSPHATE SYNTHASE.
CC SEQUENCE 406 AA; 45529 MW; OC1B9FF5EC72C1FB CRC64;
CC FT FT
CC QUERY MATCH 69.8%; Score 32; DB 1; Length 406;
CC BEST LOCAL SIMILARITY 60.0%; Pred. No. 77;
CC MATCHES 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CC QY 1 TTSETLIVE 10

```

```

Db 25 TTSETLIVE 34
RESULT 20
ISDF HELPU
ID ISDF HELPU STANDARD; PRT; 409 AA.
AC Q9ZML9;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE IspD/ispF bifunctional enzyme [Includes: 2-C-methyl-D-erythritol 4-
DE phosphate cytidyltransferase (EC 2.7.7.60) (4-diphosphocytidyl-2C-
DE methyl-D-erythritol synthase) (MEP cytidyltransferase) (MCT)]; 2-C-
DE (MECPs) (MECDP-synthase)].
GN ISPDP OR JHP0404.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]_STANDARD;
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -!- FUNCTION: Bifunctional enzyme that catalyzes the formation of 4-
CC diphosphocytidyl-2C-methyl-D-erythritol from CTP and 2C-methyl-D-
CC erythritol 4-phosphate (ispD), and converts 4-diphosphocytidyl-2C-
CC methyl-D-erythritol 2-phosphate into 2C-methyl-D-erythritol 2,4-
CC cyclodiphosphate and CMP (ispF). Also converts 4-diphosphocytidyl-
CC 2C-methyl-D-erythritol into 2C-methyl-D-erythritol 3,4-
CC cyclodiphosphate and CMP (ispF) (By similarity).
CC -!- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
CC -!- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
CC + CMP.
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ISPDP FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ISPDP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001474; AAD05981.1; -.
CC PIR; G71936; G71936.
CC DR HAMAP; MF_00107; fused; 1.
CC DR HAMAP; MF_00108; fused; 1.
CC DR InterPro; IPR001228; ISPDP synthase.
CC DR InterPro; IPR003526; YgBB.
CC DR Pfam; PF01128; ispD; 1.
CC DR Pfam; PF02542; YgBB; 1.
CC DR PIRSF; PIRSF006813; ispD/ispF synth; 1.
CC DR TIGRFAMs; TIGR00151; ispF; 1.
CC DR PROSITE; PS01295; ISPDP; 1.
CC DR PROSITE; PS01350; ISPDP; 1.
CC DR Transferase; Nucleotidyltransferase; Lyase; Isoprene biosynthesis;
CC Multifunctional enzyme; Complete proteome.
CC DOMAIN 1 247 406 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL
CC FT FT
CC QY 1 TTSETLIVE 10

```

```
FT DOMAIN 248 409 2-C-METHYL-D-ERYTHRITOL 2,4-
SQ SEQUENCE 409 AA; 45705 MW; 636B714E255DCFB0 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 409;
Best Local Similarity 60.0%; Pred. No. 78;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
DB 28 TLETLTLIK 37

RESULT 21
MURD HELPJ
ID MURD_HELPJ STANDARD; PRT; 422 AA.
AC Q9ZL70;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamate ligase (EC 6.3.2.9) (UDP-N-
acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid
adding enzyme).
DE MURD OR JHP0446.
GN Helicobacter pylori J99 (Campylobacter pylori J99).
OS Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: Cell wall formation. Catalyzes the addition of glutamate
to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanine (UMA)
(bY similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine +
glutamate = ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-
glutamate.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the murCDEF family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
CC EMBL; A8001479; A8006024.1; -.
DR F1R; C71930; C71930.
DR HSSP; P14900; 1EEH.
DR HAMAP; MF 00639; -.
DR InterPro; IPR005762; MurD.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR Pfam; PF01225; Mur_ligase; 1.
DR Pfam; PF02875; Mur_ligase_C; 1.
DR TIGRPFAMs; TIGR01087; murD; 1.
DR Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
KW ATP-binding; Complete proteome.
FT NP_BIND 102 108 ATP (POTENTIAL).
SQ SEQUENCE 422 AA; 47946 MW; 542AE103EBE22554 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 422;
```

```
Best Local Similarity 70.0%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
DB 109 TTEMLTILLE 118

RESULT 22
YDW3 SCHPO
ID YDW3_SCHPO STANDARD; PRT; 598 AA.
AC O13910;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C23C11.03 in chromosome I.
GN SPAC23C11.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
Weljens I., Vansreela E., Rieger M., Schaefer M., Mueller-Auer S.,
Babel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foraburg S.L.,
Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
Shpakovski G.V., Ussery D., Batrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: BELONGS TO THE MPPI0 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98559; CAB11156.1; -.
DR F1R; T38241; T38241..
DR GeneDB; SPombe; SPAC23C11.03; -.
DR Pfam; PF04006; Mpp10; 1.
DR Hypothetical protein.
KW Hypothetical protein.
FT DOMAIN 237 240 POLY-LYS.
FT DOMAIN 462 465 POLY-PRO.
FT DOMAIN 532 538 POLY-ARG.
SQ SEQUENCE 598 AA; 67707 MW; 36439BA7106FEF99 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 598;
```



Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTSETLIVE 10  
Db 361 TTETLEDLIK 370

## RESULT 23

MK21 YEAST  
 ID MK21 YEAST STANDARD; PRT; 1025 AA.  
 AC Q12176;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ribosome biogenesis protein MAK21.  
 GN MAK21 OR YDR060W OR D4237 OR YD9609.14.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=96381250; PubMed=8789263;  
 RA Brandt P., Ramlow S., Otto B., Bloeker H.;  
 RT "Nucleotide sequence analysis of a 32,500 bp region of the right arm  
 of Saccharomycetes cerevisiae chromosome IV.";  
 RL Yeast 12:85-90(1996).  
 [2]  
 SEQUENCE FROM N.A.  
 RX STRAIN=S288C / AB972;  
 RA Hunt S., Bowman S., Harris D., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 [3]  
 CHARACTERIZATION.  
 RX MEDLINE=99003241; PubMed=9786894;  
 RA Edskes H.K., Ontake Y., Wickner R.B.;  
 RT "Mak21p of Saccharomycetes cerevisiae, a homolog of human CNAAT-binding  
 protein, is essential for 60 S ribosomal subunit biogenesis.";  
 RL J. Biol. Chem. 273:128912-128920(1998).  
 CC -!- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE CBF / MAK21 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; X84162; CAA589976.1; -;  
 CC ENBL; Z74356; CAA98878.1; -;  
 CC ENBL; Z49209; CAA89089.1; -;  
 CC PIR; S54044; S54044.  
 CC SGD; S0002467; MAK21.  
 CC GO; GO:000027; P.ribosomal large subunit assembly and mainte. .; IMP.  
 CC InterPro; IPR005612; CBF.  
 CC Pfam; PF03914; CBF; 1.  
 CC Ribosome biogenesis; Nuclear protein.  
 CC DOMAIN 74 77 POLY-ASP.  
 CC FT DOMAIN 117 122 POLY-ASP.  
 CC FT DOMAIN 979 986 POLY-GLU.  
 CC SEQUENCE 1025 AA; 116676 MW; F880378727ED71D8 CRC64;  
 Query Match 69.6%; Score 32; DB 1; Length 1025;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 TTSETLIV 9  
 Db 302 TKSLETIV 309

RESULT 24  
 TEAL SCHPO STANDARD; PRT; 1147 AA.  
 ID TEAL SCHPO  
 AC P87061;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Tip elongation aberrant protein 1 (Cell polarity protein teal).  
 GN TEAL OR SPCC123.06.  
 OS Schizosaccharomyces pombe (Pission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=97344085; PubMed=9200612;  
 RA Mata J., Nurse P.;  
 RT "teal and the microtubular cytoskeleton are important for generating  
 global spatial order within the fission yeast cell.";  
 RL Cell 89:939-949(1997).  
 [2]  
 SEQUENCE FROM N.A.  
 RX STRAIN=972;  
 RC MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton F., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehner H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: CELL POLARITY PROTEIN. MAY ACT AS AN END MARKER,  
 CC DIRECTING THE GROWTH MACHINERY TO THE CELL POLES. MAY ALSO  
 CC INFLUENCE MICROTUBULAR ORGANIZATION, AFFECTING THE MAINTENANCE OF  
 CC A SINGLE CENTRAL AXIS.  
 CC -!- SUBCELLULAR LOCATION: PRESENT AT BOTH POLES OF THE CELL THROUGHTOUT  
 CC THE CELL CYCLE WHETHER THEY ARE GROWING OR NOT. LOCATED AT THE  
 CC ENDS OF MICROTUBULES GROWING TOWARDS THE CELL POLES. AN INTACT  
 CC MICROTUBULAR SKELETON IS REQUIRED TO MAINTAIN THE LOCATION AT THE  
 CC CELL TIPS. ON THE OTHER HAND, A NORMALLY ORGANIZED ACTIN  
 CC CYTOSKELETON IS NOT REQUIRED.  
 CC -!- SIMILARITY: Contains 5 Kelch repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC



```

DR EMBL; Y12709; CAA73246.1; -.
DR EMBL; AL031579; CAA20875.1; -.
DR PIR; T40866; T40866.
DR GenBank; SPombe; SPCC1223.06; -.
DR InterPro; IPR006652; Kelch_rep.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00612; Kelch; 3.
KW Kelch repeat; Repeat; Microtubules; Coiled coil.
FT REPEAT 94 144 KELCH 1.
FT REPEAT 146 198 KELCH 2.
FT REPEAT 254 303 KELCH 3.
FT REPEAT 305 351 KELCH 4.
FT REPEAT 355 402 KELCH 5.
FT DOMAIN 611 649 COILED COIL (POTENTIAL).
FT DOMAIN 716 838 COILED COIL (POTENTIAL).
FT DOMAIN 879 990 COILED COIL (POTENTIAL).
FT DOMAIN 1084 1105 COILED COIL (POTENTIAL).
SQ SEQUENCE 1147 AA; 127436 MW; 7BE65F6C666F4F8 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1147;
Best Local Similarity 70.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLTVE 10
   : :|||||
Db 585 TSVTLETLVE 594

RESULT 25
ACIN_MOUSE
ID ACIN_MOUSE STANDARD; PRT; 1338 AA.
DT 28-FEB-2003 (Rel. 41, Created) Q9CSX7; Q9R046; Q9R047;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Apoptotic chromatin condensation inducer in the nucleus (Acinus).
GN ACINUS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RX MEDLINE=99418558; PubMed=10490026;
RA Sahara S, Aoto M, Eguchi Y, Imanoto N, Yoneda Y, Tsujimoto Y.;
RT "Acinus is a caspase-3-activated protein required for apoptotic
chromatin condensation.";
RL Nature 401:168-173(1999).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
RP Mamoru A, Setuko S, Yoshinaka T.;
RT "Molecular cloning of murine acinusL, a gene for apoptotic chromatin
condensation.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 1-1190 FROM N.A. (ISOFORMS 1; 2 AND 4).
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Pancreas;
RX MEDLINE=22354683; PubMed=1246851;
RA Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S.,
RA Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H.,
RA Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schonbach C, Gojobori T.,
RA Baldarelli R, Hill D.P., Bult C, Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E, Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konegaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,

```

```

RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zvolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RA "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RT Nature 420:563-573(2002).
RL FUNCTION: Induces apoptotic chromatin condensation after
activation by CASP3 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=L;
CC IsoId=Q9JIX8-1; Sequence=Displayed;
CC Name=2; Synonyms=S;
CC IsoId=Q9JIX8-2; Sequence=VSP_004030, VSP_004033;
CC Name=3; Synonyms=S';
CC IsoId=Q9JIX8-3; Sequence=VSP_004031;
CC Name=4;
CC IsoId=Q9JIX8-4; Sequence=VSP_004032;
CC -!- PTM: Undergoes proteolytic cleavage; the processed form is active,
CC contrary to the uncleaved form (By similarity).
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in position 110 and 112.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF124725; AAD56723.1; -.
CC EMBL; AF124729; AAD56727.1; -.
CC EMBL; AF168782; AAF89661.1; ALT FRAME.
CC EMBL; AK011698; -. NOT ANNOTATED_CDS.
CC EMBL; AK012099; BAB28030.1; -.
CC EMBL; AK012337; BAB28171.2; -.
CC EMBL; AK050467; BAC34272.1; ALT_INIT.
CC MGD; MGI:1891824; Acinus.
CC GO; GO:0005730; C:nucleolus; IDA.
CC InterPro; IPR000504; RNA_rec_mot.
CC InterPro; IPR003034; SAP.
CC Pfam; PF02037; SAP; 1.
CC SMART; SM00513; SAP; 1.
CC PROSITE; PS50800; SAP; 1.
KW Apoptosis; Nuclear protein; Alternative splicing.
FT DOMAIN 72 106 SAP.
FT DOMAIN 142 443 GLU-RICH.
FT DOMAIN 569 667 SER-RICH.
FT DOMAIN 1131 1130 PRO-RICH.
FT DOMAIN 1131 1338 ARG/ASP/GLU/LYS-RICH.
FT SITE 1093 1094 CLEAVAGE (BY CASPASE-3) (BY SIMILARITY).
FT VARSPPLIC 1 757 Missing (in isoform 2).
FT VARSPPLIC 1 773 Missing (in isoform 3).
FT VARSPPLIC 164 204 Missing (in isoform 4).
FT VARSPPLIC 758 766 Missing (in isoform 2).
FT CONFLICT 244 244 Missing (in isoform 2).
FT CONFLICT 515 515 T -> A (IN REF. 2).

```

```

FT CONFLICT 536 536 F -> L (IN REF. 2).
FT CONFLICT 557 557 G -> D (IN REF. 2).
FT CONFLICT 568 568 H -> Y (IN REF. 2).
FT CONFLICT 599 599 V -> A (IN REF. 2).
FT CONFLICT 729 729 S -> I (IN REF. 2).
FT CONFLICT 757 757 L -> P (IN REF. 2).
FT CONFLICT 773 773 T -> A (IN REF. 1; AAD56723).
FT CONFLICT 829 829 MISSING (IN REF. 3; BAB28030).
FT CONFLICT 896 896 Q -> R (IN REF. 3; BAB28030).
FT CONFLICT 1035 1035 G -> R (IN REF. 3; BAB28030).
SQ SEQUENCE 1338 AA; 150690 MW; B912D9CB5750BFB4 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1338;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETIV 9
    |||||:
Db 867 TTESLKSLI 875

RESULT 26
ACIN HUMAN
ID ACIN HUMAN STANDARD; PRT; 1341 AA.
AC Q9UKV3; O75158; Q9UG91; Q9UKV1; Q9UKV2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apoptotic chromatin condensation inducer in the nucleus (Acinus).
GN ACINUS OR KIAA0670.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), PARTIAL SEQUENCE, FUNCTION,
RP AND MUTAGENESIS OF ASP-1093.
RX MEDLINE=99418558; PubMed=10490026;
RA Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;
RT "Acinus is a caspase-3-activated protein required for apoptotic
chromatin condensation.";
RL Nature 401:168-173(1999).
[2]
RN SEQUENCE FROM N.A. (ISOFORM 4).
RP TISSUE=Uterus;
RC Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=fetal brain;
RA Li W.B., Gruber C., Jesse J., Polayes D.;
RT "Full-length cDNA libraries and normalization.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 56-1341 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
CC -!- FUNCTION: Induces apoptotic chromatin condensation after
activation by CASP3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=L;
CC IsoId=Q9UKV3-1; Sequence=Displayed;
CC Name=2; Synonyms=S;
CC IsoId=Q9UKV3-2; Sequence=VSP_004025, VSP_004028;
CC Name=3; Synonyms=S;

CC CONFLICT 536 536 F -> L (IN REF. 2).
CC CONFLICT 557 557 G -> D (IN REF. 2).
CC CONFLICT 568 568 H -> Y (IN REF. 2).
CC CONFLICT 599 599 V -> A (IN REF. 2).
CC CONFLICT 729 729 S -> I (IN REF. 2).
CC CONFLICT 757 757 L -> P (IN REF. 2).
CC CONFLICT 773 773 T -> A (IN REF. 1; AAD56723).
CC CONFLICT 829 829 MISSING (IN REF. 3; BAB28030).
CC CONFLICT 896 896 Q -> R (IN REF. 3; BAB28030).
CC CONFLICT 1035 1035 G -> R (IN REF. 3; BAB28030).
SQ SEQUENCE 1338 AA; 150690 MW; B912D9CB5750BFB4 CRC64;

Name=4;
IsoId=Q9UKV3-4; Sequence=VSP_004026, VSP_004029;
-!- TISSUE SPECIFICITY: Ubiquitous
-!- PTM: Undergoes proteolytic cleavage; the processed form is active,
contrary to the uncleaved form.
-!- SIMILARITY: Contains 1 SAP domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; AF124726; AAD56724.1; -
CC EMBL; AF124727; AAD56725.1; -
CC EMBL; AF124728; AAD56726.1; -
CC EMBL; AL050382; CAB43681.1; -
CC EMBL; BX247975; CAD62309.1; -
CC EMBL; AB014570; BAA31645.2; -
CC MIM; 604562; -
CC InterPro; IPR000504; RNA_rec_mot.
CC InterPro; IPR003034; SAP.
CC Pfam; PF02037; SAP; 1.
CC SMART; SM00513; SAP; 1.
CC PROSITE; PS50800; SAP; 1.
KW Apoptosis; Nuclear protein; Alternative splicing.
FT DOMAIN 72 106 SAP.
FT DOMAIN 142 442 GLU-RICH.
FT DOMAIN 573 676 SER-RICH.
FT DOMAIN 1114 1131 PRO-RICH.
FT DOMAIN 1132 1341 ARG/ASP/GLU/LYS-RICH.
FT SITE 1093 1094 CLEAVAGE (BY CASPASE-3).
FT VARSPPLIC 1 727 Missing (in isoform 2).
FT VARSPPLIC 1 758 Missing (in isoform 3).
FT VARSPPLIC 1 1152 Missing (in isoform 4).
FT VARSPPLIC 728 766 GSPKCEAEAEPPAATOPQTSQTSHLPSEIRHHV
P -> MSPADRCASANTIEPATTSLLALFULLORDQSSRRGL
(in isoform 2).
FT VARSPPLIC 759 766 /FTID=VSP_004028.
SERIHHV -> MLSEKSG (in isoform 3).
FT MUTAGEN 1093 1093 D->A; ABOLISHES CLEAVAGE BY CASP3 AND
CHROMATIN CONDENSATION ACTIVITY.
FT CONFLICT 139 139 Q -> H (IN REF. 4).
SQ SEQUENCE 1341 AA; 151887 MW; 8FE286681F83AB5C CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1341;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETIV 9
    |||||:
Db 867 TTESLKSLI 875

RESULT 27
RPOB COXBU
ID RPOB COXBU STANDARD; PRT; 1375 AA.
AC O87903;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
beta chain) (RNA polymerase beta subunit).
GN RPOB OR CBU0231.
OS Coxiella burnetii.

```



```

RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -!- PATHWAY: Fatty acid biosynthesis pathway; first reduction step.
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
CC (SDR) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32701; AAC21824.1; -.
CC PIR: D64051; D64051.
CC HSSP: P19992; LHDC.
CC TIGR: H10155; -.
CC InterPro: IPR002198; ADH short.
CC Pfam: PF00106; adh short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 8 32 NADP (BY SIMILARITY).
FT ACT_SITE 149 149 BY SIMILARITY.
SQ SEQUENCE 242 AA; 25507 MW; B3DBE2C020D2F71 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 242;
Best Local Similarity 75.0%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10
DB 62 ESLETLVE 69

RESULT 30
Y638 RHIL0 STANDARD; PRT; 367 AA.
AC Q98MC1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease ml10638 (EC 3.4.24.-).
GN ML10638.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFE303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

```

```

RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
CC -!- COFACTOR: Zinc (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP002995; BAB48192.1; ALT_INIT.
CC MEROPS: M50.004; -.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001193; Peptidase M50.
CC InterPro: IPR004387; Zn_Mprotease.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF02163; Peptidase_M50; 1.
CC SMART: SM00228; PDZ; 1.
CC TIGRFAMs: TIGR00054; TIGR00054; 1.
CC PROSITE: PS0106; PDZ; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW Inner membrane; Complete proteome.
FT METAL 20 20 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 21 21 POTENTIAL.
FT METAL 24 24 ZINC (CATALYTIC) (POTENTIAL).
FT TRANSMEM 108 130 POTENTIAL.
FT TRANSMEM 291 313 POTENTIAL.
FT TRANSMEM 343 365 POTENTIAL.
FT DOMAIN 121 196 PDZ.
SQ SEQUENCE 367 AA; 39922 MW; 0E77DF0CB96D5152 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 367;
Best Local Similarity 60.0%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTSELETLVE 10
DB 82 TTSELETLVE 91

RESULT 31
DDC ACIBA STANDARD; PRT; 510 AA.
AC Q43908;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-2,4-diaminobutyrate decarboxylase (EC 4.1.1.-) (DABA decarboxylase)
DE (DABA-DC).
GN DDC.
OS Acinetobacter baumannii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=470;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 19606 / NCTC 12156 / CIP 70.34;
RX MEDLINE=96337867; PubMed=8772175;
RA Ikai H., Yamamoto S.;
RT "Sequence analysis of the gene encoding a novel L-2,4-diaminobutyrate
RT decarboxylase of Acinetobacter baumannii: similarity to the group II
RT amino acid decarboxylases.";
RL Arch. Microbiol. 166:128-131(1996).
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: 1,3-diaminopropane biosynthesis.

```

```
CC -!- SIMILARITY: BELONGS TO THE GROUP II DECARBOXYLASE FAMILY (DDC,
CC GAD, HDC AND TYRDC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D55724; BAA09538.1; -.
CC InterPro; IPR002129; Pyridoxal dec.
CC Pfam; PF00282; Pyridoxal dec; 1.
CC PROSITE; PS00392; DDC GAD HDC YDC; 1.
CC Lysase; Decarboxylase; Pyridoxal phosphate.
KW BINDING 319 319 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 510 AA; 56244 MW; 900DF52FD1941B70 CRC64;
Query Match 67.4%; Score 31; DB 1; Length 510;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TESLETLVE 10
Db 410 TEGLELLVE 418
-----
RESULT 32
SEST CAEL STANDARD; PRT; 517 AA.
ID SEST CAEL
AC Q9N4D6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative sestrin.
GN Y74C9A.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RA Du H.; Wohldmann P.; Ames M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RW CONCEPTUAL TRANSLATION.
RP Axelisen K.; Bairoch A.;
RL Unpublished observations (MAR-2001).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the sestrin family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC024206; AAF36051.1; ALT_SEQ.
CC WormPep; Y74C9A.5; CE24663.
CC InterPro; IPR006730; PA26.
CC Pfam; PF04636; PA26; 1.
KW Hypothetical protein; Nuclear protein.
SQ SEQUENCE 517 AA; 60881 MW; D7BC041916D0E205 CRC64;
Query Match 67.4%; Score 31; DB 1; Length 517;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TTESLETLV 9
Db 294 TTETIESLI 302
-----
RESULT 33
YGSU YEAST STANDARD; PRT; 577 AA.
ID YGSU YEAST
AC P53333;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 67.3 kDa protein in TAF145-YOR1 intergenic region.
GN YGR278W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=S288c / FY1679;
RX MEDLINE=97245295; PubMed=9090054;
RA Volckaert G.; Voet M.; Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MAL1 locus reveals 15 complete open reading frames, including
RT ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";
RL Yeast 13:251-259(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73063; CAA97309.1; -.
CC FIR; S64613; S64613.
CC SGD; S0003510; YGR278W.
CC InterPro; IPR003890; IF_eIF4G.
CC InterPro; IPR003891; IF_eIF4G_MA3.
CC Pfam; PF02847; MA3; 1.
CC Pfam; PF02854; MIF4G; 1.
CC SMART; SM00544; MIF4G; 1.
CC SMART; SM00543; MIF4G; 1.
KW Hypothetical protein.
SQ SEQUENCE 577 AA; 67294 MW; 773E33C71E4A29D0 CRC64;
Query Match 67.4%; Score 31; DB 1; Length 577;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ESLETLVE 10
Db 192 ESLETLFE 199
-----
RESULT 34
PRIM CHLTR STANDARD; PRT; 595 AA.
ID PRIM CHLTR
AC O84799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA primase (EC 2.7.7.-).
GN DNAG OR CT794.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RC SEQUENCE FROM N.A.
```

```
RC STRAIN=D/UW-3/Cx; PubMed:9784136;
RX MEDLINE=9900809;
RA Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: Monomer (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001351; AAC68389.1; -
CC PIR; F71471; F71471.
CC HSSP; Q9X4D0; 1DQ0.
CC InterPro; IPR006295; DNA_primase.
CC InterPro; IPR006171; Toprim_dom.
CC InterPro; IPR006647; Toprim_primase.
CC InterPro; IPR006154; Toprim_sub.
CC InterPro; IPR002694; Znf_CHC2.
CC Pfam; PF01751; Toprim; 1.
CC Pfam; PF01807; zf-CHC2; 1.
CC ProDom; PD002276; Toprim_primase; 1.
CC ProDom; PD002988; Znf_CHC2; 1.
CC SMART; SM00493; TOPRIM; 1.
CC SMART; SM00400; Znf_CHCC; 1.
CC TIGRfams; TIGR01391; dnaG; 1.
KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
FT ZNF_FING 38 62 CHC2-TYPE (BY SIMILARITY).
SQ SEQUENCE 595 AA; 68037 MW; 536858BBAFCDF8FB6 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 595;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8
DB 4 TTESLETL 11

RESULT 35
ID YF64 CAEEL STANDARD; PRT; 598 AA.
AC Q21890;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein R102.4 in chromosome IV.
GN R102.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: BELONGS TO THE THREONINE ALDOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002284; AAF39049.1; -
CC PIR; H81733; H81733.
CC HSSP; Q9X4D0; 1DQ0.
CC TIGR; TC0175; -
CC InterPro; IPR006295; DNA_primase.
CC InterPro; IPR006171; Toprim_dom.
CC InterPro; IPR006647; Toprim_primase.
CC InterPro; IPR006154; Toprim_sub.
CC InterPro; IPR002694; Znf_CHC2.
CC Pfam; PF01751; Toprim; 1.
CC Pfam; PF01807; zf-CHC2; 1.
CC ProDom; PD002276; Toprim_primase; 1.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 270309; CAA94358.1; -
CC PIR; D88817; D88817.
CC WormPep; R102.4; CE06332.
KW Hypothetical protein; Lyase; Pyridoxal phosphate.
FT BINDING 255 255 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 598 AA; 67009 MW; EDAL670E82181D26 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 598;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10
DB 387 ENLETLVE 394

RESULT 36
ID PRIM CHLMU STANDARD; PRT; 600 AA.
AC Q9PLC9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA primase (EC 2.7.7.-).
GN DNAG OR TC0175.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -!- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: Monomer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002284; AAF39049.1; -
CC PIR; H81733; H81733.
CC HSSP; Q9X4D0; 1DQ0.
CC TIGR; TC0175; -
CC InterPro; IPR006295; DNA_primase.
CC InterPro; IPR006171; Toprim_dom.
CC InterPro; IPR006647; Toprim_primase.
CC InterPro; IPR006154; Toprim_sub.
CC InterPro; IPR002694; Znf_CHC2.
CC Pfam; PF01751; Toprim; 1.
CC Pfam; PF01807; zf-CHC2; 1.
CC ProDom; PD002276; Toprim_primase; 1.
```

```

DR ProDom; PD002988; Znf_CHC2; 1.
DR SMART; SM00493; TOPRIM; 1.
DR SMART; SM00400; Znf_CHCC; 1.
DR TIGRFAMs; TIGR01391; dnaG; 1.
KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;
KW Zinc-finger; Zinc; Metal-binding; Complete proteome.
FT ZN_FING 38 62 CHC2-TYPE (BY SIMILARITY).
SQ SEQUENCE 600 AA; 68810 MW; 040C08EB754FF41A CRC64;

Query Match 67.4%; Score 31; DB 1; Length 600;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETL 8
Db 4 TTESLETL 11

RESULT 37
V162 FOWPV STANDARD; PRT; 603 AA.
AC Q9J569;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ankryrin-repeat protein FPV162.
GN FPV162.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
ON NCBI_TaxID=10261;
RN SEQUENCE FROM N.A.
RP MEDLINE=20193820; PubMed=10729156;
RX Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831(2000).
CC -!- SIMILARITY: Contains 14 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF198100; AAF44506.1; -.
DR HSSP; P25963; I1KN.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 12.
DR SMART; SM00248; ANK; 13.
DR PROSITE; PS00888; ANK_REPEAT; 11.
DR PROSITE; PS02997; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; Repeat; ANK repeat.
FT REPEAT 23 53
FT REPEAT 57 87
FT REPEAT 91 120
FT REPEAT 124 155
FT REPEAT 159 189
FT REPEAT 193 223
FT REPEAT 227 257
FT REPEAT 261 291
FT REPEAT 295 325
FT REPEAT 329 362
FT REPEAT 366 397
FT REPEAT 401 428
FT REPEAT 432 467
FT REPEAT 504 533
SQ SEQUENCE 603 AA; 66542 MW; 8F8E3E464DB9D16D CRC64;

Query Match 67.4%; Score 31; DB 1; Length 603;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETL 10
Db 173 TNKTITLIE 182

RESULT 38
GHR RABIT STANDARD; PRT; 638 AA.
AC P19341;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Growth hormone receptor precursor (GH receptor) (Serum binding
DE protein).
DE
GN GHR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_TaxID=9986;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88065896; PubMed=2825030;
RA Leung D.W., Spencer S.A., Cachianes G., Hammonds R.G., Collins C.,
RA Henzel W.J., Barnard R., Waters M.J., Wood W.I.;
RT "Growth hormone receptor and serum binding protein: purification,
RT cloning and expression";
RL Nature 330:537-543(1987).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR PITUITARY GLAND GROWTH HORMONE.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF015252; AAB67613.1; -.
DR PIR; S08544; B28176.
DR HSSP; P10912; IA22.
DR InterPro; IPR002996; CR1A.
DR InterPro; IPR003961; FN.III.
DR Pfam; PF00041; fn3; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS01352; HEMATOPO_REC_L_F1; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 638
FT DOMAIN 19 264
FT TRANSMEM 265 288
FT DOMAIN 289 338
FT DOMAIN 145 252
FT DISULFID 56 66
FT DISULFID 101 112
FT DISULFID 126 140
FT CARBOHYD 46 46
FT CARBOHYD 115 115
FT CARBOHYD 156 156
FT CARBOHYD 161 161
FT CARBOHYD 200 200
SQ SEQUENCE 638 AA; 71076 MW; E05CCE1D7294624C CRC64;

Query Match 67.4%; Score 31; DB 1; Length 638;
Best Local Similarity 70.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 TTSETLV 10
Db 568 TTSETLV 577

RESULT 39
SUW8_ARATH STANDARD; PRT; 755 AA.
ID SUW8_ARATH STANDARD; PRT; 755 AA.
AC O9C5F0; Q9TP24;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 8
DE (c 2.1.1.43) (Histone H3-K9 methyltransferase 8) (H3-K9-HMTase 8)
DE (Suppressor of variegation 3-9 homolog 8) (Su(var)3-9 homolog 8)
GN SUW8 OR AT2G24740 OR F27A10.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RN RP MEDLINE=21550130; PubMed=11691919;
RX RA Baumbach L.O., Thorstensen T., Krauss V., Fischer A., Naumann K.,
RA Assalkhou R., Schulz I., Reuter G., Aalen R.B.;
RT "The Arabidopsis thaliana genome contains at least 29 active genes
RT encoding SET domain proteins that can be assigned to four
RT evolutionarily conserved classes.";
RN [2] Nucleic Acids Res. 29:4319-4333(2001).
RN RP
RN RP SEQUENCE FROM N.A.
RN RP STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Roundley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
CC -!- FUNCTION: Histone methyltransferase. Methylates Lys-9 of histone
CC H3. H3 Lys-9 methylation represents a specific tag for epigenetic
CC transcriptional repression.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -!- SUBCELLULAR LOCATION: Nuclear; associates with centromeric
CC constitutive heterochromatin (By similarity).
CC -!- DOMAIN: Although the SET domain contains the active site of
CC enzymatic activity, both pre-SET and post-SET domains are required
CC for methyltransferase activity.
CC -!- SIMILARITY: BELONGS TO THE HISTONE-LYSINE METHYLTRANSFERASE
CC FAMILY. SUVAR3-9 SUBFAMILY.
CC -!- SIMILARITY: Contains 1 A/T hook DNA-binding repeat.
CC -!- SIMILARITY: Contains 1 YDG domain.
CC -!- SIMILARITY: Contains 1 pre-SET domain.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF344451; AAK28973.1; -.
CC DR EMBL; AC007266; AAD26896.1; ALT_INIT.
CC DR EMBL; X82775; CAA58017.1; -.
CC DR EMBL; Z72983; CAA97225.1; -.
CC DR PIR; S53921; S53921.
CC DR SGD; S0003430; YGR198W.
CC DR YHypothetical protein.

DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR003105; G9A.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR007728; Pre-SET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR003606; Zn2-binding.
DR Pfam; PF02178; AT_hook; 1.
DR Pfam; PF05033; Pre-SET; 1.
DR Pfam; PF00856; SET; 1.
DR Pfam; PF02182; YDG_SRA; 1.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00468; PreSET; 1.
DR SMART; SM00317; SET; 1.
DR SMART; SM00466; SRA; 1.
DR PROSITE; PS00868; POST_SET; 1.
DR PROSITE; PS00867; PRE_SET; FALSE_NEG.
DR PROSITE; PS0280; SET; 1.
KW Transfrase; Methyltransferase; Chromatin regulator; Nuclear protein;
KW DNA-binding; Multigene family.
FT DNA_BIND 174 186 A.T HOOK.
FT DOMAIN 305 452 YDG.
FT DOMAIN 528 578 PRE-SET.
FT DOMAIN 580 727 SET.
FT DOMAIN 739 755 POST-SET.
SQ SEQUENCE 755 AA; 84527 MW; 4B5BF379B8BF0C27 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 755;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTSETLV 9
Db 368 TTSETLV 375

RESULT 40
YH4B_YEAST STANDARD; PRT; 817 AA.
ID YH4B_YEAST STANDARD; PRT; 817 AA.
AC P46951;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 95.4 kDa protein in SNG1-PMT6 intergenic region.
GN YGR198W OR G7594.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RN RP STRAIN=S288C;
RX MEDLINE=96076633; PubMed=7502584;
RA Guerreiro P., Maia e Silva A., Batreiros T., Atroyo J.,
RA Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C.,
RA Nombela C.;
RT "The complete sequence of a 9000 bp fragment of the right arm of
RT Saccharomyces cerevisiae chromosome VII contains four previously
RT unknown open reading frames.";
RL Yeast 11:1087-1091(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X82775; CAA58017.1; -.
CC DR EMBL; Z72983; CAA97225.1; -.
CC DR PIR; S53921; S53921.
CC DR SGD; S0003430; YGR198W.
CC DR YHypothetical protein.

```



SQ .SEQUENCE 817 AA; 95365 MW; AC42730C8B9C3E4B CRC64;

Query Match

Best Local Similarity 67.4%; Score 31; DB 1; Length 817;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTESLETIVE 10

Db 590 TLEALETPE 599

Search completed: October 30, 2003, 14:14:45  
Job time : 26 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2003, 14:11:46 ; Search time 28 Seconds  
(without alignments)  
15.111 Million cell updates/sec

Title: US-09-868-293B-2  
Perfect score: 46  
Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 46    | 100.0       | 132    | 4  | US-09-198-452A-96    |
| 2          | 33    | 71.7        | 154    | 4  | US-09-491-916-2      |
| 3          | 33    | 71.7        | 157    | 4  | US-09-134-001C-5103  |
| 4          | 33    | 71.7        | 429    | 4  | US-09-252-991A-28983 |
| 5          | 33    | 71.7        | 500    | 1  | US-08-117-083-68     |
| 6          | 33    | 71.7        | 599    | 4  | US-09-328-352-4814   |
| 7          | 32    | 69.6        | 274    | 4  | US-09-134-001C-5279  |
| 8          | 32    | 69.6        | 307    | 2  | US-08-807-861A-56    |
| 9          | 32    | 69.6        | 307    | 3  | US-09-210-681-56     |
| 10         | 32    | 69.6        | 307    | 3  | US-08-946-719A-56    |
| 11         | 32    | 69.6        | 307    | 4  | US-09-547-983-56     |
| 12         | 32    | 69.6        | 567    | 4  | US-09-205-258-573    |
| 13         | 31    | 67.4        | 242    | 4  | US-09-107-532A-3774  |
| 14         | 31    | 67.4        | 283    | 4  | US-09-107-532A-6481  |
| 15         | 31    | 67.4        | 487    | 4  | US-09-328-352-4816   |
| 16         | 31    | 67.4        | 525    | 4  | US-09-328-352-4492   |
| 17         | 31    | 67.4        | 577    | 3  | US-09-315-794-32     |
| 18         | 31    | 67.4        | 577    | 3  | US-09-389-341-32     |
| 19         | 31    | 67.4        | 620    | 3  | US-09-000-145-3      |
| 20         | 31    | 67.4        | 778    | 4  | US-09-252-991A-30592 |
| 21         | 31    | 67.4        | 3144   | 1  | US-08-246-982A-6     |
| 22         | 31    | 67.4        | 3144   | 2  | US-08-453-265-6      |
| 23         | 31    | 67.4        | 3144   | 2  | US-08-457-273B-42    |
| 24         | 31    | 67.4        | 3144   | 3  | US-08-556-419-21     |
| 25         | 31    | 67.4        | 3144   | 3  | US-09-041-886-15     |
| 26         | 30    | 65.2        | 30     | 2  | US-08-248-839C-80    |
| 27         | 30    | 65.2        | 191    | 4  | US-09-252-991A-17100 |

|    |    |      |      |   |                      |                    |
|----|----|------|------|---|----------------------|--------------------|
| 28 | 30 | 65.2 | 289  | 3 | US-08-961-083-20     | Sequence 20, Appl  |
| 29 | 30 | 65.2 | 289  | 4 | US-09-536-784-20     | Sequence 20, Appl  |
| 30 | 30 | 65.2 | 294  | 4 | US-09-107-532A-6796  | Sequence 6796, Ap  |
| 31 | 30 | 65.2 | 309  | 2 | US-08-715-131-2      | Sequence 2, Appli  |
| 32 | 30 | 65.2 | 309  | 3 | US-08-221-733-2      | Sequence 2, Appli  |
| 33 | 30 | 65.2 | 315  | 4 | US-09-134-001C-3527  | Sequence 3527, Ap  |
| 34 | 30 | 65.2 | 328  | 4 | US-09-428-747-4      | Sequence 4, Appli  |
| 35 | 30 | 65.2 | 484  | 1 | US-08-030-096-8      | Sequence 8, Appli  |
| 36 | 30 | 65.2 | 580  | 4 | US-09-134-001C-5611  | Sequence 5611, Ap  |
| 37 | 30 | 65.2 | 624  | 4 | US-09-252-991A-21625 | Sequence 21625, A  |
| 38 | 30 | 65.2 | 953  | 2 | US-08-506-340A-1     | Sequence 1, Appli  |
| 39 | 30 | 65.2 | 983  | 4 | US-09-394-200-2      | Sequence 2, Appli  |
| 40 | 30 | 65.2 | 3696 | 4 | US-09-134-001C-5080  | Sequence 5080, Ap  |
| 41 | 29 | 63.0 | 105  | 4 | US-09-130-158A-4     | Sequence 4, Appli  |
| 42 | 29 | 63.0 | 120  | 4 | US-08-178-257-11     | Sequence 11, Appli |
| 43 | 29 | 63.0 | 120  | 6 | 5218100-6            | Patent No. 5218100 |
| 44 | 29 | 63.0 | 122  | 4 | US-09-107-532A-6701  | Sequence 6701, Ap  |
| 45 | 29 | 63.0 | 132  | 4 | US-09-107-532A-6694  | Sequence 6694, Ap  |

ALIGNMENTS

RESULT 1  
US-09-198-452A-96  
; Sequence 96, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Grifais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 96  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-96

Query Match 100.0%; Score 46; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
DB 5 TTESLETLVE 14

RESULT 2  
US-09-491-916-2  
; Sequence 2, Application US/09491916  
; Patent No. 6326462  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; TITLE OF INVENTION: No. 6326462el ribh  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/491,916

```

; FILING DATE: 27-Jan-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/977,553
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-491-916-2

Query Match          71.7%; Score 33; DB 4; Length 154;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
Db 116 TTESIEQAVE 125

RESULT 3
US-09-134-001C-5103
; Sequence 5103, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5103
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5103

Query Match          71.7%; Score 33; DB 4; Length 157;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10
Db 120 TTESIEQAVE 129

RESULT 4
US-09-252-991A-28983
; Sequence 28983, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28983
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28983

Query Match          71.7%; Score 33; DB 4; Length 429;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTESLETLVE 10
Db 281 TEALEVLVE 289

RESULT 5
US-08-117-083-68
; Sequence 68, Application US/08117083
; Patent No. 5719054
; GENERAL INFORMATION:
; APPLICANT: Boursnell, Michael E.
; APPLICANT: Inglis, Stephen C.
; APPLICANT: Munro, Alan J.
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
; TITLE OF INVENTION: Papilloma Virus Proteins
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Walter H. Dreger
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,083
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-58783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..500
; OTHER INFORMATION: /note="Xaa refers to stop codon in
; OTHER INFORMATION: the open reading frame."
US-08-117-083-68

Query Match          71.7%; Score 33; DB 1; Length 500;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

Qy 2 TESLETLVE 10  
 |||:||||  
 Db 110 TESIERLVE 118

## RESULT 6

US-09-328-352-4814  
 ; Sequence 4814, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 4814  
 ; LENGTH: 599  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-4814

Query Match 71.7%; Score 33; DB 4; Length 599;  
 Best Local Similarity 75.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ESLETLVE 10  
 |||:||||  
 Db 444 ESLETLII 451

## RESULT 7

US-09-134-001C-5279  
 ; Sequence 5279, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5279  
 ; LENGTH: 274  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5279

Query Match 69.6%; Score 32; DB 4; Length 274;  
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTLESLETLV 9  
 |||:||||  
 Db 250 TTKSLEALV 258

## RESULT 8

US-08-807-861A-56  
 ; Sequence 56, Application US/08807861A  
 ; Patent No. 5853975  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tartaglia, Louis A.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
 ; FILE OF INVENTION: REGULATION OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/807,861A  
 FILING DATE: 26-FEB-1997  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/518,878  
 FILING DATE: 23-AUG-1995  
 APPLICATION NUMBER: US 08/470,868  
 FILING DATE: 06-JUN-1995  
 APPLICATION NUMBER: US 08/294,522  
 FILING DATE: 23-AUG-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7853-066  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 56:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 307 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-807-861A-56

Query Match 69.6%; Score 32; DB 2; Length 307;  
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTLESLETL 8  
 |||:||||  
 Db 166 TTESLSLTL 173

## RESULT 9

US-09-210-681-56  
 ; Sequence 56, Application US/09210681  
 ; Patent No. 6057109  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tartaglia, Louis A.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
 ; FILE OF INVENTION: REGULATION OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/210,681  
 FILING DATE:  
 CLASSIFICATION:

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,861
; FILING DATE: 26-FEB-1997
; APPLICATION NUMBER: US 08/518,878
; FILING DATE: 23-AUG-1995
; APPLICATION NUMBER: US 08/470,868
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/294,522
; FILING DATE: 23-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 307 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-210-681-56

Query Match 69.6%; Score 32; DB 3; Length 307;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8
Db 166 TTESLSTL 173

RESULT 10
US-08-946-719A-56
; Sequence 56, Application US/08946719A
; Patent No. 6121017
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,719A
; FILING DATE: 8-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/807,861
; FILING DATE: 26-FEB-1997
; APPLICATION NUMBER: US 08/518,878
; FILING DATE: 23-AUG-1995
; APPLICATION NUMBER: US 08/470,868
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/294,522
; FILING DATE: 23-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-107
; US-09-868-293b-2.ra1
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 307 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-946-719A-56

Query Match 69.6%; Score 32; DB 3; Length 307;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8
Db 166 TTESLSTL 173

RESULT 11
US-09-547-983-56
; Sequence 56, Application US/09547983
; Patent No. 6518402
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: REGULATION OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/547,983
; FILING DATE: 12-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/807,861
; FILING DATE: 26-FEB-1997
; APPLICATION NUMBER: US 08/518,878
; FILING DATE: 23-AUG-1995
; APPLICATION NUMBER: US 08/470,868
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: US 08/294,522
; FILING DATE: 23-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 307 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
; US-09-547-983-56
; US-09-868-293b-2.ra1
```

Query Match 69.6%; Score 32; DB 4; Length 307;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETL 8  
Db 166 TTESLSTL 173

RESULT 12  
US-09-205-258-573  
; Sequence 573, Application US/09205258  
; Patent No. 6525174  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258  
; CURRENT FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,915  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,019  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,970  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,972  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,916  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,373  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,875  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,374

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,917  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,949  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,883  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,897  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,898  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,962  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 573  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (409)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-205-258-573

Query Match 69.6%; Score 32; DB 4; Length 567;  
Best Local Similarity 66.7%; Pred. No. 2.9e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTESLETLV 9  
Db 94 TTESLKSLLI 102

RESULT 13  
US-09-107-532A-3774  
; Sequence 3774, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:

QY 2 TESLETLVE 10



```
Db 425 TEGLELLVE 433

RESULT 17
; Sequence 32, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-053
; CURRENT APPLICATION NUMBER: US/09/315,794
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-315-794-32

Query Match 67.4%; Score 31; DB 3; Length 577;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ESLETLVE 10
Db 192 ESLETLFE 199

RESULT 18
US-09-389-341-32
; Sequence 32, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-057
; CURRENT APPLICATION NUMBER: US/09/389,341
; CURRENT FILING DATE: 1999-09-02
; EARLIER APPLICATION NUMBER: 09/315,794
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-389-341-32

Query Match 67.4%; Score 31; DB 3; Length 577;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ESLETLVE 10
Db 192 ESLETLFE 199

RESULT 19
US-09-000-145-3
; Sequence 3, Application US/09000145
; Patent No. 6169172
; GENERAL INFORMATION:
; APPLICANT: DEVAUCHELLE, Gerrard
; APPLICANT: GARNIER, Laurence
; APPLICANT: CAHOREAU, Claire
; APPLICANT: CERUTTI, Martine
; TITLE OF INVENTION: USE OF A PROLACTIN RECEPTOR OR GROWTH HORMONE RECEPTOR
```

```
; TITLE OF INVENTION: INTRACYTOPLASMIC DOMAIN FOR ACHIEVING PROTEIN SECRETION
; FILE REFERENCE: 0384-0047-0XPCT
; CURRENT APPLICATION NUMBER: US/09/000,145
; CURRENT FILING DATE: 1998-03-16
; EARLIER APPLICATION NUMBER: PCT/FR96/01237
; EARLIER FILING DATE: 1996-08-02
; EARLIER APPLICATION NUMBER: FR 95/09420
; EARLIER FILING DATE: 1995-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-000-145-3

Query Match 67.4%; Score 31; DB 3; Length 620;
Best Local Similarity 70.0%; Pred. No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTESLETLVE 10
Db 550 TTESLTAAE 559

RESULT 20
US-09-252-991A-30592
; Sequence 30592, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30592
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30592

Query Match 67.4%; Score 31; DB 4; Length 778;
Best Local Similarity 66.7%; Pred. No. 6.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10
Db 722 TESLEALME 730

RESULT 21
US-08-246-982A-6
; Sequence 6, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
;
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,982A
; FILING DATE: May 20, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge, A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3880002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-246-982A-6

Query Match 67.4%; Score 31; DB 1; Length 3144;
Best Local Similarity 77.8%; Pred. No. 2.9e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10
Db 868 TELLETTLAE 876

RESULT 22
US-08-453-265-6
; Sequence 6, Application US/08453265
; Patent No. 5693757
; GENERAL INFORMATION:
; APPLICANT: Macdonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James P.
; TITLE OF INVENTION: Huntington DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,265
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3880003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
;
; MOLECULE TYPE: protein
; US-08-453-265-6

Query Match 67.4%; Score 31; DB 1; Length 3144;
Best Local Similarity 77.8%; Pred. No. 2.9e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10
Db 868 TELLETTLAE 876

RESULT 23
US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; RELATED DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-457-273B-42

Query Match 67.4%; Score 31; DB 2; Length 3144;
Best Local Similarity 77.8%; Pred. No. 2.9e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10
Db 868 TELLETTLAE 876

RESULT 24
US-08-556-419-21
; Sequence 21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
```

APPLICANT: Lanahan, Anthony  
APPLICANT: Worley, Paul  
APPLICANT: Snyder, Solomon  
TITLE OF INVENTION: Huntingtin-associated protein  
FILE REFERENCE: 01107.52271  
CURRENT APPLICATION NUMBER: US/08/556,419C  
CURRENT FILING DATE: 1995-11-09  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 21  
LENGTH: 3144  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-08-556-419-21

Query Match 67.4%; Score 31; DB 3; Length 3144;  
Best Local Similarity 77.8%; Pred. No. 2.9e+03;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10  
||| ||| |||  
Db 868 TELLETAE 876

RESULT 25  
US-09-041-886-15  
Sequence 15, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabinzadeh, Shiroz  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041.886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31.815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3144 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-041-886-15

Query Match 67.4%; Score 31; DB 3; Length 3144;  
Best Local Similarity 77.8%; Pred. No. 2.9e+03;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TESLETLVE 10  
||| ||| |||  
Db 868 TELLETAE 876

RESULT 26  
US-08-248-839C-80  
Sequence 80, Application US/08248839C  
Patent No. 5843702  
GENERAL INFORMATION:  
APPLICANT: McConnell, David  
APPLICANT: Devine, Kevin  
APPLICANT: O'Kane, Charles  
TITLE OF INVENTION: A Gene Expression System  
NUMBER OF SEQUENCES: 185  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/248.839C  
FILING DATE: 25-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER: 3614.214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-248-839C-80

Query Match 65.2%; Score 30; DB 2; Length 30;  
Best Local Similarity 75.0%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TESLETLV 9  
||| ||| |||  
Db 16 TESLRTLI 23

RESULT 27  
US-09-252-991A-17100  
Sequence 17100, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17100  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17100



REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6796:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 294 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...294  
SEQUENCE DESCRIPTION: SEQ ID NO: 6796:  
US-09-107-532A-6796

Query Match 65.2%; Score 30; DB 4; Length 294;  
Best Local Similarity 75.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
DB 153 TVESLETL 160

RESULT 31  
US-08-715-131-2  
Sequence 2, Application US/08715131  
Patent No. 5854416  
GENERAL INFORMATION:  
APPLICANT: Sampson, Jacquelyn S.  
APPLICANT: Russell, Harold  
APPLICANT: Tharpe, Jean A.  
APPLICANT: Ades, Edwin W.  
APPLICANT: Carlone, George M.  
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37-KDA SURFACE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Needle & Rosenberg, P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303-1811  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/715,131  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 14114.0200  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 309 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-715-131-2

Query Match 65.2%; Score 30; DB 2; Length 309;  
Best Local Similarity 60.0%; Pred. No. 3.5e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
DB 232 TPEQIKTLVE 241

RESULT 32  
US-09-221-753-2  
Sequence 2, Application US/09221753  
Patent No. 6217884  
GENERAL INFORMATION:  
APPLICANT: SAMPSON, JACQUELYN S.  
APPLICANT: RUSSELL, HAROLD  
APPLICANT: THARPE, JEAN A.  
APPLICANT: ADES, EDWIN W.  
APPLICANT: CARLONE, GEORGE M.  
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37 KDA SURFACE  
FILE REFERENCE: 64778 US  
CURRENT APPLICATION NUMBER: US/09/221,753  
CURRENT FILING DATE: 1998-12-28  
EARLIER APPLICATION NUMBER: US 07/791,377  
EARLIER FILING DATE: 1991-09-17  
EARLIER APPLICATION NUMBER: US 07/816,286  
EARLIER FILING DATE: 1992-01-03  
EARLIER APPLICATION NUMBER: US 08/222,179  
EARLIER FILING DATE: 1994-04-04  
EARLIER APPLICATION NUMBER: US 08/715,131  
EARLIER FILING DATE: 1996-09-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 309  
TYPE: PRT  
ORGANISM: STREPTOCOCCUS PNEUMONIAE  
US-09-221-753-2

Query Match 65.2%; Score 30; DB 3; Length 309;  
Best Local Similarity 60.0%; Pred. No. 3.5e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
DB 232 TPEQIKTLVE 241

RESULT 33  
US-09-134-001C-3527  
Sequence 3527, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3527  
LENGTH: 315  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3527

Query Match 65.2%; Score 30; DB 4; Length 315;  
Best Local Similarity 75.0%; Pred. No. 3.6e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
|:|:|:|:|  
Db 176 ESIKTLVE 183

RESULT 34

US-09-428-747-4  
; Sequence 4, Application US/09428747  
; Patent No. 6387681  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Shuang-yong  
; APPLICANT: Xiao, Jian-ping  
; TITLE OF INVENTION: Method For Cloning And Expression Of NheI Restriction  
; TITLE OF INVENTION: Endonuclease In E. coli  
; FILE REFERENCE: NheI  
; CURRENT APPLICATION NUMBER: US/09/428,747  
; CURRENT FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Neisseria mucosa heidelbergensis  
US-09-428-747-4

Query Match 65.2%; Score 30; DB 4; Length 328;  
Best Local Similarity 62.5%; Pred. No. 3.7e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
|:|:|:|:|  
Db 49 DSLETLIE 56

RESULT 35

US-08-030-096-8  
; Sequence 8, Application US/08030096  
; Patent No. 5426041  
; GENERAL INFORMATION:  
; APPLICANT: Fabijanski, Steven F.  
; APPLICANT: Arnison, Paul G.  
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID  
; TITLE OF INVENTION: SEED PRODUCTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/030,096  
; FILING DATE: 22-MAR-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/556,917  
; FILING DATE: 20-JUL-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA91/00255  
; FILING DATE: 22-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 33229/164/PIHI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 484 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-030-096-8

Query Match 65.2%; Score 30; DB 1; Length 484;  
Best Local Similarity 75.0%; Pred. No. 5.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
|:|:|:|:|  
Db 124 EDLETLIVE 131

RESULT 36

US-09-134-001C-5611  
; Sequence 5611, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5611  
; LENGTH: 580  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5611

Query Match 65.2%; Score 30; DB 4; Length 580;  
Best Local Similarity 75.0%; Pred. No. 6.9e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
|:|:|:|:|  
Db 518 EALETIME 525

RESULT 37

US-09-252-991A-21625  
; Sequence 21625, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21625  
; LENGTH: 624  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21625

Query Match 65.2%; Score 30; DB 4; Length 624;

Best Local Similarity 75.0%; Pred. No. 7.5e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 0;

Qy 3 ESLETLVE 10  
Db 558 ESLETLVQ 565

RESULT 38  
US-08-506-340A-1  
; Sequence 1, Application US/08506340A  
; Patent No. 5846810  
; GENERAL INFORMATION:  
; APPLICANT: Yano, Keichi  
; APPLICANT: Yamagaki, Motoo  
; APPLICANT: Tanaka, Keiji  
; TITLE OF INVENTION: HUMAN 26S PROTEASOME SUBUNIT COMPONENTS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO  
; STREET: 277 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10172-0194  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/506.340A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 264810/94  
; FILING DATE: 28-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lawrence S. Perry  
; REGISTRATION NUMBER: 31865  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-758-2400  
; TELEFAX: 212-758-2982  
; TELEX: 236262  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 953 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; CELL TYPE: human hepatic carcinoma cell strain HepG2

Query Match 65.2%; Score 30; DB 2; Length 953;  
Best Local Similarity 60.0%; Pred. No. 1.2e+03;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTSELTIVE 10  
Db 917 TSIEDIEELVE 926

RESULT 39  
US-09-394-200-2  
; Sequence 2, Application US/09394200  
; Patent No. 6353154  
; GENERAL INFORMATION:  
; APPLICANT: Kossmann, Jens  
; APPLICANT: Froberg, Claus  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; STARCH PHOSPHORYLASE FROM MAIZE

NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10020-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/394.200  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP98/01183  
FILING DATE: 03-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 197 09 775.8  
FILING DATE: 10-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: PTB-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 983 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-394-200-2

Query Match 65.2%; Score 30; DB 4; Length 983;  
Best Local Similarity 75.0%; Pred. No. 1.2e+03;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ESLETLVE 10  
Db 524 KSLETLVE 531

RESULT 40  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5080  
; LENGTH: 3696  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5080

Query Match 65.2%; Score 30; DB 4; Length 3696;  
Best Local Similarity 60.0%; Pred. No. 5.2e+03;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTSELTIVE 10

Db        655 TTEKOTLVD 664  
          ||| :|||:

Search completed: October 30, 2003, 14:18:30  
Job time : 30 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: October 30, 2003, 14:18:01 ; Search time 70 Seconds  
(without alignments)  
24.449 Million cell updates/sec

Title: US-09-868-293B-2  
Perfect score: 46  
Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 642050 seqs, 171146064 residues

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 36    | 78.3        | 2993   | 10 | US-09-738-626-6239  |
| 2          | 35    | 76.1        | 264    | 15 | US-10-156-761-10082 |
| 3          | 34    | 73.9        | 122    | 9  | US-09-734-569-54    |
| 4          | 34    | 73.9        | 305    | 9  | US-09-734-569-158   |
| 5          | 33    | 71.7        | 539    | 15 | US-10-106-698-6386  |
| 6          | 33    | 71.7        | 585    | 9  | US-09-922-217-1067  |
| 7          | 33    | 71.7        | 585    | 10 | US-09-833-263-1067  |
| 8          | 33    | 71.7        | 585    | 14 | US-10-025-380-1067  |
| 9          | 32    | 69.6        | 85     | 12 | US-10-190-435-310   |
| 10         | 32    | 69.6        | 92     | 10 | US-09-764-877-1507  |
| 11         | 32    | 69.6        | 567    | 12 | US-09-933-767-573   |
| 12         | 32    | 69.6        | 567    | 15 | US-10-023-282-573   |
| 13         | 31    | 67.4        | 18     | 12 | US-10-029-386-32825 |
| 14         | 31    | 67.4        | 241    | 10 | US-09-978-248-11    |
| 15         | 31    | 67.4        | 242    | 9  | US-09-815-242-10974 |

Sequence 156, App  
Sequence 5479, App  
Sequence 12149, A  
Sequence 34039, A  
Sequence 4312, App  
Sequence 7, Appli  
Sequence 33924, A  
Sequence 20, Appl  
Sequence 156, App  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 7700, App  
Sequence 348, App  
Sequence 11001, A  
Sequence 11073, A  
Sequence 5766, App  
Sequence 7122, App  
Sequence 3608, App  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 42481, A  
Sequence 33589, A  
Sequence 34157, A  
Sequence 30184, A  
Sequence 34035, A  
Sequence 5880, App  
Sequence 5604, App  
Sequence 12176, A

## ALIGNMENTS

### RESULT 1

```
US-09-738-626-6239
; Sequence 6239, Application US/09738626
; Publication No. US00020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6239
; LENGTH: 2993
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6239
```

Query Match 78.3%; Score 36; DB 10; Length 2993;  
Best Local Similarity 77.8%; Pred. No. 8.3e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTESLETLVE 10

Db 1740 TDSIETLVE 1748

RESULT 2

US-10-156-761-10082

; Sequence 10082, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 10082

; LENGTH: 264

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-10082

Query Match 76.1%; Score 35; DB 15; Length 264;

Best Local Similarity 77.8%; Pred. No. 85;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TDSIETLVE 10

Db 22 TEALDTLVE 30

RESULT 3

US-09-734-569-54

; Sequence 54, Application US/09734569

; Patent No. US20020064816A1

; GENERAL INFORMATION:

; APPLICANT: Lerchl, Jens

; APPLICANT: Renz, Andreas

; APPLICANT: Ehrhardt, Thomas

; APPLICANT: Reindl, Andreas

; APPLICANT: Cirpus, Petra

; APPLICANT: Bischoff, Friedrich

; APPLICANT: Frank, Markus

; APPLICANT: Freund, Annette

; APPLICANT: Duwenig, Elke

; APPLICANT: Schmidt, Ralf-Michael

; APPLICANT: Reski, Ralf

; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved

; FILE REFERENCE: BASF-NAE-1332-99-US

; CURRENT APPLICATION NUMBER: US/09/734,569

; CURRENT FILING DATE: 2001-05-24

; PRIOR APPLICATION NUMBER: US 60/171,101

; PRIOR FILING DATE: 1999-12-16

; NUMBER OF SEQ ID NOS: 181

; SOFTWARE: Patentin Ver. 2.1/WordPerfect 6.1

; SEQ ID NO 54

; LENGTH: 122

; TYPE: PRT

; ORGANISM: Physcomitrella patens

US-09-734-569-54

Query Match 73.9%; Score 34; DB 9; Length 122;

Best Local Similarity 70.0%; Pred. No. 55;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

Db 67 TTESIKKLVE 76

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

Db 72 TTESIKKLVE 81

RESULT 4

US-09-734-569-158

; Sequence 158, Application US/09734569

; Patent No. US20020064816A1

; GENERAL INFORMATION:

; APPLICANT: Lerchl, Jens

; APPLICANT: Renz, Andreas

; APPLICANT: Ehrhardt, Thomas

; APPLICANT: Reindl, Andreas

; APPLICANT: Cirpus, Petra

; APPLICANT: Bischoff, Friedrich

; APPLICANT: Frank, Markus

; APPLICANT: Freund, Annette

; APPLICANT: Duwenig, Elke

; APPLICANT: Schmidt, Ralf-Michael

; APPLICANT: Reski, Ralf

; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved

; FILE REFERENCE: BASF-NAE-1332-99-US

; CURRENT APPLICATION NUMBER: US/09/734,569

; CURRENT FILING DATE: 2001-05-24

; PRIOR APPLICATION NUMBER: US 60/171,101

; PRIOR FILING DATE: 1999-12-16

; NUMBER OF SEQ ID NOS: 181

; SOFTWARE: Patentin Ver. 2.1/WordPerfect 6.1

; SEQ ID NO 158

; LENGTH: 305

; TYPE: PRT

; ORGANISM: Physcomitrella patens

US-09-734-569-158

Query Match 73.9%; Score 34; DB 9; Length 305;

Best Local Similarity 70.0%; Pred. No. 1.5e+02;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

Db 67 TTESIKKLVE 76

RESULT 5

US-10-106-698-6386

; Sequence 6386, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

; FILE REFERENCE: PA005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: Patentin Ver. 3.0

; SEQ ID NO 6386

; LENGTH: 539

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: MISC\_FEATURE

; LOCATION: (2)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: MISC\_FEATURE  
; LOCATION: (7)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (8)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (11)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-106-698-6386

Query Match 71.7%; Score 33; DB 15; Length 539;  
Best Local Similarity 87.5%; Pred. No. 4.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||  
Db 180 STESLETL 187

## RESULT 6

US-09-922-217-1067  
; Sequence 1067, Application US/09922217  
; Patent No. US20020076414A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole Lynn  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.471C13  
; CURRENT APPLICATION NUMBER: US/09/922,217  
; CURRENT FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1067  
; LENGTH: 585  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-217-1067

Query Match 71.7%; Score 33; DB 9; Length 585;  
Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||  
Db 226 STESLETL 233

## RESULT 7

US-09-833-263-1067  
; Sequence 1067, Application US/09833263  
; Patent No. US20020110547A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1067  
; LENGTH: 585  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-263-1067

Query Match 71.7%; Score 33; DB 10; Length 585;  
Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||  
Db 226 STESLETL 233

## RESULT 8

US-10-025-380-1067  
; Sequence 1067, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1067  
; LENGTH: 585  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-025-380-1067

Query Match 71.7%; Score 33; DB 14; Length 585;  
Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||  
Db 226 STESLETL 233

## RESULT 9

US-10-190-435-310  
; Sequence 310, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Esclrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP18133.003 / 2302-18133

```

; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 310
;   LENGTH: 85
;   TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: vpu 92BR025
US-10-190-435-310

Query Match          69.6%; Score 32; DB 12; Length 85;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 TESLETLVE 10
DB      66 TTEETMVD 74

RESULT 10
US-09-764-877-1507
; Sequence 1507, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1507
;   LENGTH: 92
;   TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1507

Query Match          69.6%; Score 32; DB 10; Length 92;
Best Local Similarity 60.0%; Pred. No. 94;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTSELTVE 10
DB      54 TTSEIQAE 63

RESULT 11
US-09-933-767-573
; Sequence 573, Application US/09933767
; Publication No. US20030181692A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P2
; CURRENT APPLICATION NUMBER: US/09/933,767
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: PCT/US01/05614
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,836
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/193,170
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 09/205,258
; PRIOR FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: PCT/US98/11422
; PRIOR FILING DATE: 1998-06-04

```

; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/068,064  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/068,053  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/070,923  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/073,160  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/073,159  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/073,165  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/073,164  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/085,925  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,921  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,923  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,922  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/092,921  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/094,657  
; PRIOR FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1245  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 573  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (409)  
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids

US-09-933-767-573

Query Match 69.6%; Score 32; DB 12; Length 567;

Best Local Similarity 66.7%; Pred. No. 7.2e+02;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTESLETLV 9

Db 94 TTESLSLI 102

RESULT 12

US-10-023-282-573

; Sequence 573, Application US/10023282

; Publication No. US20030092893A1

; GENERAL INFORMATION:

; APPLICANT: Young et al.

; TITLE OF INVENTION: 207 Human Secreted Proteins

; FILE REFERENCE: PZ007P1

; CURRENT APPLICATION NUMBER: US/10/023,282

; CURRENT FILING DATE: 2001-12-20

; EARLIER APPLICATION NUMBER: 09/205,258

; EARLIER FILING DATE: 1998-12-04

; EARLIER APPLICATION NUMBER: PCT/US98/11422

; EARLIER FILING DATE: 1998-06-04

; EARLIER APPLICATION NUMBER: 60/048,885

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,375

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,881

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,880

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/048,896

; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,020

```
;
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (409)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-023-282-573

Query Match      69.6%; Score 32; DB 15; Length 567;
Best Local Similarity 66.7%; Pred. No. 7.2e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTSETLTV 9
Db 94 TTESLSLI 102

RESULT 13
US-10-029-386-32825
; Sequence 32825, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AEOMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32825
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL050303.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
US-10-029-386-32825

Query Match      67.4%; Score 31; DB 12; Length 18;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTSETLTV 9
Db 4 TTEELEAI 12

RESULT 14
US-09-978-248-11
; Sequence 11, Application US/09978248
; Publication No. US20020197690A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Pamela M
; APPLICANT: Virca, Duke G
; APPLICANT: Bird, Timothy A
; APPLICANT: Garka, Kristen
; TITLE OF INVENTION: GID (GNK INTERACTING DECARBOXYLASE) AND METHODS OF USE
; FILE REFERENCE: 2499-1-001N
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: US/09/978,248
; PRIOR FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,324
; PRIOR FILING DATE: 2001-10-18
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-978-248-11
```

```
Query Match      67.4%; Score 31; DB 10; Length 241;
Best Local Similarity 77.8%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TSETLTV 10
Db 221 TEGLELVE 229

RESULT 15
US-09-815-242-10974
; Sequence 10974, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10974
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10974

Query Match      67.4%; Score 31; DB 9; Length 242;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 BSETLTV 10
Db 62 ESIETLLE 69

RESULT 16
US-09-734-569-156
; Sequence 156, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
```

APPLICANT: Schmidt, Ralf-Michael  
APPLICANT: Reski, Ralf  
TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved  
FILE REFERENCE: BASP-NAE-1332-99-US  
CURRENT APPLICATION NUMBER: US/09/734,569  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 60/171,101  
PRIOR FILING DATE: 1999-12-16  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1  
SEQ ID NO 156  
LENGTH: 327  
TYPE: PRT  
ORGANISM: Physcomitrella patens  
US-09-734-569-156

Query Match 67.4%; Score 31; DB 9; Length 327;  
Best Local Similarity 60.0%; Pred. No. 6e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETIVE 10  
Db 88 TTESIKLVLD 97

RESULT 17  
US-09-815-242-5479  
Sequence 5479, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011a  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5479  
LENGTH: 449  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5479

Query Match 67.4%; Score 31; DB 9; Length 449;  
Best Local Similarity 66.7%; Pred. No. 8.5e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETIV 9  
Db 106 TKETLETLI 114

RESULT 18  
US-09-815-242-12149  
Sequence 12149, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011a  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12149  
LENGTH: 452  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12149

Query Match 67.4%; Score 31; DB 9; Length 452;  
Best Local Similarity 66.7%; Pred. No. 8.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETIV 9  
Db 109 TKETLETLI 117

RESULT 19  
US-10-029-386-34039  
Sequence 34039, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34039  
LENGTH: 669  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-029-386-34039

Query Match 66.7%; Score 31; DB 9; Length 449;  
Best Local Similarity 66.7%; Pred. No. 8.5e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
; OTHER INFORMATION: SWISSPROT HIT: P17221, EVALUE 2.00e-71
US-10-029-386-34039

Query Match      67.4%; Score 31; DB 12; Length 669;
Best Local Similarity 66.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTSETLV 9
Db 356 TTELEALI 364

RESULT 20
US-09-738-626-4312
; Sequence 4312, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 4312
; LENGTH: 1233
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4312

Query Match      67.4%; Score 31; DB 10; Length 1233;
Best Local Similarity 77.8%; Pred. No. 2.6e+03;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLV 9
Db 214 TTESUKQV 222

RESULT 21
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. US20020037908A1
; GENERAL INFORMATION:
; APPLICANT: NO. US20020037908Alactyl, Inc.
; TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and Prepath
; TITLE OF INVENTION: Protein Assembly or Aggregation
; FILE REFERENCE: 42108/26146
; CURRENT APPLICATION NUMBER: US/09/904,987
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 7
; LENGTH: 1543
; TYPE: PRT
; ORGANISM: homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
```

```
; DATABASE ENTRY DATE: 2001-04-16
; RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

Query Match      67.4%; Score 31; DB 9; Length 1543;
Best Local Similarity 77.8%; Pred. No. 3.4e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TESLETLV 10
Db 865 TELLETLAE 873

RESULT 22
US-10-029-386-33924
; Sequence 33924, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33924
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009516.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.26
; OTHER INFORMATION: SWISSPROT HIT: Q53192, EVALUE 3.90e+00
US-10-029-386-33924

Query Match      65.2%; Score 30; DB 12; Length 183;
Best Local Similarity 75.0%; Pred. No. 4.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLV 10
Db 60 DSLETLV 67

RESULT 23
US-09-765-272-20
; Sequence 20, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
```



```

; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 289 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-765-272-20

```

```

Query Match          65.2%; Score 30; DB 9; Length 289;
Best Local Similarity 60.0%; Pred. No. 8e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TTSETLTVE 10
    ||::|||
DB 212 TPEQIKTLVE 221

```

```

RESULT 24
US-09-769-787-156
; Sequence 156, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansdro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 156
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-156

```

```

Query Match          65.2%; Score 30; DB 11; Length 309;
Best Local Similarity 60.0%; Pred. No. 8.6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TTSETLTVE 10
    ||::|||
DB 232 TPEQIKTLVE 241

```

```

RESULT 25
US-09-754-809-2
; Sequence 2, Application US/09754809
; Publication No. US20030105307A1
; GENERAL INFORMATION:
; APPLICANT: SAMPSON, JACQUELYN S.
; APPLICANT: RUSSELL, HAROLD
; APPLICANT: THARPE, JEAN A.
; APPLICANT: ADES, EDWIN W.
; APPLICANT: CARLONE, GEORGE M.
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37 kDa SURFACE
; TITLE OF INVENTION: ADHESIN A PROTEIN
; FILE REFERENCE: 64778 US

```

```

; CURRENT APPLICATION NUMBER: US/09/754,809
; CURRENT FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: US/09/221,753
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 07/791,377
; PRIOR FILING DATE: EARLIER FILING DATE: 1991-09-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 07/816,286
; PRIOR FILING DATE: EARLIER FILING DATE: 1992-01-03
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 08/222,179
; PRIOR FILING DATE: EARLIER FILING DATE: 1994-04-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 08/715,131
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 309
; TYPE: PRT
; ORGANISM: STREPTOCOCCUS PNEUMONIAE
US-09-754-809-2

```

```

Query Match          65.2%; Score 30; DB 11; Length 309;
Best Local Similarity 60.0%; Pred. No. 8.6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TTSETLTVE 10
    ||::|||
DB 232 TPEQIKTLVE 241

```

```

RESULT 26
US-09-893-519A-2
; Sequence 2, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 75
US-09-893-519A-2

```

```

Query Match          65.2%; Score 30; DB 11; Length 330;
Best Local Similarity 66.7%; Pred. No. 9.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 TESLETLVE 10
      || :|||:|
Db      167 TEFIEU 175

RESULT 27
US-10-032-585-7700
; Sequence 7700, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7700
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7700

Query Match      65.2%; Score 30; DB 12; Length 330;
Best Local Similarity 66.7%; Pred. No. 9.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 TESLETLVE 10
      || :|||:|
Db      167 TEFIEU 175

RESULT 28
US-10-177-293-348
; Sequence 348, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05

```

```

; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 348
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-348

Query Match      65.2%; Score 30; DB 15; Length 372;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 TESLETLVE 10
      |||||:|
Db      314 TESLESLTQ 322

RESULT 29
US-10-156-761-11001
; Sequence 11001, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11001
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11001

Query Match      65.2%; Score 30; DB 15; Length 398;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 TESLETLVE 10
      |||||:|
Db      170 SOSLETLTE 178

RESULT 30
US-09-815-242-11073
; Sequence 11073, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

```

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11073  
; LENGTH: 456  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-815-242-11073

Query Match 65.2%; Score 30; DB 9; Length 456;  
Best Local Similarity 60.0%; Pred. No. 1.3e+03;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLV 10  
| : : : : :  
Db 110 TKETLEKLE 119

RESULT 31  
US-09-738-626-5766  
; Sequence 5766, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5766  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5766

Query Match 65.2%; Score 30; DB 10; Length 618;  
Best Local Similarity 66.7%; Pred. No. 1.9e+03;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLV 9  
| : : : : :  
Db 513 TTSEEIALV 521

RESULT 32  
US-10-032-585-7122  
; Sequence 7122, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7122  
; LENGTH: 841  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7122

Query Match 65.2%; Score 30; DB 12; Length 841;  
Best Local Similarity 60.0%; Pred. No. 2.6e+03;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLV 10  
| : : : : :  
Db 533 TSSVESLVE 542

RESULT 33  
US-09-738-626-3608  
; Sequence 3608, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3608  
; LENGTH: 925  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3608

Query Match 65.2%; Score 30; DB 10; Length 925;  
Best Local Similarity 75.0%; Pred. No. 2.9e+03;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
| : : : : :  
Db 792 STEALETL 799

RESULT 34

us-09-868-293b-2.rapb

Thu Oct 30 14:52:59 2003

```

; SEQ ID NO 6
; LENGTH: 1734
; TYPE: PRT
; ORGANISM: Spodoptera frugiperda
; US-10-126-339-6

Query Match      65.2%; Score 30; DB 12; Length 1734;
Best Local Similarity 75.0%; Pred. No. 5.9e+03; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 0;

Qy 3 ESLETLVE 10
Db 1418 ESLQTVVE 1425

RESULT 36
US-10-126-148-6
; Sequence 6, Application US/10126148
; Publication No. US20030170787A1
; GENERAL INFORMATION:
; APPLICANT: Flannagan, Ronald D.
; APPLICANT: Mathis, John P.
; APPLICANT: Meyer, Terry E.
; TITLE OF INVENTION: No. US20030170787A1e1 Bt Toxin Receptors From
; FILE REFERENCE: 35718/204664
; CURRENT APPLICATION NUMBER: US/10/126,148
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/715,909
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/166,285
; PRIOR FILING DATE: 1999-11-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1734
; TYPE: PRT
; ORGANISM: Spodoptera frugiperda
; US-10-126-148-6

Query Match      65.2%; Score 30; DB 12; Length 1734;
Best Local Similarity 75.0%; Pred. No. 5.9e+03; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 0;

Qy 3 ESLETLVE 10
Db 1418 ESLQTVVE 1425

RESULT 37
US-09-920-653-3
; Sequence 3, Application US/09920653
; Patent No. US20020104113A1
; GENERAL INFORMATION:
; APPLICANT: Japan as Represented by Director General of Okazaki National Research
; APPLICANT: Institutes
; TITLE OF INVENTION: Nav2 channel gene-deficient non-human animals
; FILE REFERENCE: U2001P059
; CURRENT APPLICATION NUMBER: US/09/920,653
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: JP 2000/237320
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: JP 2000/241637
; PRIOR FILING DATE: 2000-08-09
; PRIOR APPLICATION NUMBER: JP 2001/222263
; PRIOR FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1681
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-920-653-3

```

```
Query Match      64.1%; Score 29.5; DB 10; Length 1681;
Best Local Similarity 80.0%; Pred. No. 7.1e+03;
Matches      8; Conservative      1; Mismatches      0; Indels      1; Gaps      1;

QY      1 TTSETLIVE 10
      ||||| :||
DB      14 TTSETLIVE 22

RESULT 38
US-09-864-761-42481
; Sequence 42481, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42481
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011609.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
```

```
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: SWISSPROT HIT: P35331, EVALUE 1.00e-10
; OTHER INFORMATION: EST_HUMAN HIT: BF342999.1, EVALUE 2.00e-09
US-09-864-761-42481

Query Match      63.0%; Score 29; DB 9; Length 55;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches      6; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

QY      2 TTSETLIVE 10
      |||| :|||
DB      5 TTSETLIVE 13

RESULT 39
US-09-864-761-33589
; Sequence 33589, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33589
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO AP000142.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 22
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 20
US-09-864-761-33589
```

```
Query Match 63.0%; Score 29; DB 9; Length 67;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 TTESLETLY 9
Db 58 TTESVEEV 66
```

## RESULT 40

```
US-09-864-761-34157
; Sequence 34157, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34157
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000089.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 8.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.2
US-09-864-761-34157
```

```
Query Match 63.0%; Score 29; DB 9; Length 67;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 TTESLETLY 9
Db 58 TTESVEEV 66
```

```
Search completed: October 30, 2003, 14:28:00
Job time : 72 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2003, 14:03:06 ; Search time 83 Seconds  
(without alignments)  
19.124 Million cell updates/sec

Title: US-09-868-293B-2  
Perfect score: 46  
Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 46    | 100.0       | 10     | 21 AAB13403 | Chlamydia trachoma |
| 2          | 46    | 100.0       | 129    | 22 AAG63380 | Amino acid sequenc |
| 3          | 46    | 100.0       | 130    | 23 ABB08370 | Ribosomal protein  |
| 4          | 46    | 100.0       | 132    | 20 AAY34678 | Chlamydia pneumoni |
| 5          | 37    | 80.4        | 88     | 22 ABG16153 | Novel human diagno |
| 6          | 37    | 80.4        | 581    | 22 ABB63559 | Drosophila melanog |
| 7          | 36    | 78.3        | 202    | 21 AAG38900 | Arabidopsis thalia |
| 8          | 36    | 78.3        | 218    | 21 AAG35722 | Arabidopsis thalia |
| 9          | 36    | 78.3        | 272    | 21 AAG38899 | Arabidopsis thalia |

|    |    |      |      |    |          |                    |
|----|----|------|------|----|----------|--------------------|
| 10 | 36 | 78.3 | 288  | 21 | AAG35721 | Arabidopsis thalia |
| 11 | 36 | 78.3 | 300  | 21 | AAG38898 | Arabidopsis thalia |
| 12 | 36 | 78.3 | 316  | 21 | AAG35720 | Arabidopsis thalia |
| 13 | 36 | 78.3 | 1277 | 22 | AAB76536 | Corynebacterium gl |
| 14 | 36 | 78.3 | 2012 | 22 | AAB76534 | Corynebacterium gl |
| 15 | 36 | 78.3 | 2993 | 22 | AAG92485 | C glutamicum prote |
| 16 | 34 | 73.9 | 122  | 22 | AAM00048 | Plastidic trioseph |
| 17 | 34 | 73.9 | 305  | 22 | AAM00100 | Plastidic trioseph |
| 18 | 33 | 71.7 | 68   | 22 | ABBI6242 | Human nervous syst |
| 19 | 33 | 71.7 | 153  | 22 | AAG83035 | S. epidermidis ope |
| 20 | 33 | 71.7 | 154  | 20 | AAY17249 | S. aureus ribH pol |
| 21 | 33 | 71.7 | 154  | 21 | AAB10801 | Staphylococcus aur |
| 22 | 33 | 71.7 | 157  | 23 | ABP40258 | Staphylococcus epi |
| 23 | 33 | 71.7 | 501  | 13 | AAR27741 | Sequence transcrib |
| 24 | 33 | 71.7 | 530  | 23 | ABB90301 | Human polypeptide  |
| 25 | 33 | 71.7 | 539  | 22 | AAM92426 | Human digestive sy |
| 26 | 33 | 71.7 | 539  | 22 | AAG75612 | Human colon cancer |
| 27 | 33 | 71.7 | 585  | 21 | AAY59290 | Human MUC12 polype |
| 28 | 33 | 71.7 | 585  | 22 | AAM24515 | C903P predicted am |
| 29 | 33 | 71.7 | 585  | 24 | ABP55364 | Human colon tumour |
| 30 | 33 | 71.7 | 601  | 23 | ABB97588 | Novel human protei |
| 31 | 33 | 71.7 | 615  | 22 | AAM79256 | Human protein SFQ  |
| 32 | 33 | 71.7 | 2440 | 18 | AAW20828 | H. pylori cytoplas |
| 33 | 32 | 69.6 | 92   | 22 | ABB03560 | Human musculoskele |
| 34 | 32 | 69.6 | 92   | 24 | ABU12854 | Novel human muscul |
| 35 | 32 | 69.6 | 125  | 18 | AAW28030 | Staphylococcus aur |
| 36 | 32 | 69.6 | 156  | 21 | AAB10806 | Shigella flexneri  |
| 37 | 32 | 69.6 | 156  | 21 | AAB10807 | Escherichia coli 1 |
| 38 | 32 | 69.6 | 156  | 22 | AAB70801 | E. coli ribE prote |
| 39 | 32 | 69.6 | 164  | 22 | AAO05824 | Human polypeptide  |
| 40 | 32 | 69.6 | 169  | 6  | AAP50078 | Sequence of the C- |
| 41 | 32 | 69.6 | 189  | 22 | ABG25591 | Novel human diagno |
| 42 | 32 | 69.6 | 189  | 22 | ABG28423 | Novel human diagno |
| 43 | 32 | 69.6 | 216  | 22 | AAG82785 | S. epidermidis ope |
| 44 | 32 | 69.6 | 216  | 24 | ABU19037 | Pathogen specific  |
| 45 | 32 | 69.6 | 236  | 24 | ABU19104 | Pathogen specific  |

ALIGNMENTS

RESULT 1

|          |    |  |
|----------|----|--|
| AAB13403 | ID | AAB13403 standard; peptide; 10 AA.                                     |
| XX       | AC | AAB13403;  |
| XX       | DT | 14-NOV-2000 (first entry)  |
| XX       | DE | Chlamydia trachomatis ribosomal protein L7/L12 N-terminus.             |
| XX       | KW | Chlamydia trachomatis; L2 strain; ribosomal protein L7/L12; immunogen; |
| XX       | KW | trachoma; infectious pneumonitis; sexually transmitted infection;      |
| XX       | KW | immunoblot; antigen.   |
| XX       | OS | Chlamydia trachomatis.   |
| XX       | PN | WO200037494-A2.  |
| XX       | PD | 29-JUN-2000.   |
| XX       | PF | 17-DEC-1999; 99WO-IB02065.   |
| XX       | PR | 18-DEC-1998; 98GB-0028000.   |
| XX       | PA | (CHIR-) CHIRON SPA.  |
| XX       | PI | Ratti G;   |
| XX       | DR | WPI; 2000-442639/38.   |
| XX       | PT | New Chlamydia trachomatis protein useful in the treatment, diagnosis   |
| PT       |    | and prevention of Chlamydia infections -                               |

```

XX PS Claim 2; Page 16; 25pp; English.
XX CC The present sequence is the N-terminus of the ribosomal protein L7/L12
XX CC from the L2 strain of Chlamydia trachomatis. Immunoblot analysis showed
XX CC that the protein is immunogenic in humans as a consequence of chlamydial
XX CC infection. A number of Chlamydia trachomatis immunogens were identified
XX CC by this method. The immunogenic proteins may be used in the manufacture
XX CC of medicaments for the treatment and prevention of chlamydial infection.
XX CC Antibodies against the proteins and nucleic acids encoding the proteins
XX CC may be used in the diagnosis of infection. The Chlamydia are responsible
XX CC for endemic sexually transmitted infections, trachoma, infectious
XX CC pneumonitis, and various other disease syndromes.
XX SQ Sequence 10 AA;

Query Match 100.0%; Score 46; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.028; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Oy 1 TTESLETLVE 10
Db 1 TTESLETLVE 10
|||||

RESULT 2
AAG63380
ID AAG63380 standard; Protein; 129 AA.
XX AC AAG63380;
XX DT 15-OCT-2001 (first entry)
XX DE Amino acid sequence of a ribosomal protein L7/L12.
XX KW Ribosomal protein; L7/L12 protein; pneumonia.
XX OS Chlamydia pneumoniae.
XX PN WO200157089-A1.
XX PD 09-AUG-2001.
XX PF 31-JAN-2001; 2001WO-JP00625.
XX PR 31-JAN-2000; 2000JP-0062684.
XX PA (ASAH ) ASAH KASEI KOGYO KK.
XX PI Rahman M, Etoh T;
XX DR WPI; 2001-483420/52.
XX DR N-PSDB; AAH74226.
XX PT Detection of Chlamydia pneumoniae by using antibody against ribosomal
XX PT protein, applicable in diagnosis of (non-stereotypic) pneumonia by
XX PT assaying the causative microbe e.g. in body fluids like blood and
XX PT tissue samples -
XX PS Example 1; Page 26; 30pp; Japanese.
XX CC The present sequence represents a ribosomal protein, designated L7/L12
XX CC protein. The specification describes an antibody that reacts
XX CC specifically with the ribosomal protein of Chlamydia pneumoniae,
XX CC particularly L7/L12 protein. The antibody is used for the detection
XX CC of Chlamydia pneumoniae, which is applicable in diagnosis of
XX CC (non-stereotypic) pneumonia by assaying the causative microbe e.g. in
XX CC body fluids like blood and tissue samples, particularly useful in
XX CC pharmaceutical industry and clinical medicine.
XX SQ Sequence 129 AA;

Query Match 100.0%; Score 46; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 0.048; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Oy 1 TTESLETLVE 10
Db 2 TTESLETLVE 11
|||||

Query Match 100.0%; Score 46; DB 23; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.49; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Oy 1 TTESLETLVE 10
Db 2 TTESLETLVE 11
|||||

RESULT 4
AAV34678
ID AAV34678 standard; Protein; 132 AA.
XX AC AAV34678;
XX DT 13-SEP-1999 (first entry)
XX DE Chlamydia pneumoniae transmembrane protein sequence.
XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX KW vaccine; neutralising epitope.
XX OS Chlamydia pneumoniae.

```



|          |  |
|----------|--|
| Pt       | New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -   |
| Cc       | The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. |
| Cc       | Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.   |
| Sq       | Sequence 88 AA;  |
| Qy       | Query Match 80.4%; Score 37; DB 22; Length 88;<br>Best Local Similarity 80.0%; Pred. No. 15;<br>Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;   |
| Dd       | 1 TTESLETLVE 10<br>       <br>39 TEESLDTLVE 48   |
| Result 6 |  |
| Abg63559 |  |
| ID       | ABG63559 standard; Protein; 581 AA.  |
| AC       | ABB63559;  |
| XX       |  |
| XX       |  |
| DT       | 26-MAR-2002 (first entry)  |
| XX       |  |
| DE       | Drosophila melanogaster polypeptide SEQ ID NO 17469.   |
| XX       |  |
| KW       | Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.   |
| XX       |  |
| OS       | Drosophila melanogaster.   |
| XX       |  |
| PN       | WO200171042-A2.  |
| XX       |  |
| PD       | 27-SEP-2001.   |
| XX       |  |
| PF       | 23-MAR-2001; 2001WO-US09231.   |
| XX       |  |
| PR       | 23-MAR-2000; 2000US-191637P.   |
| PR       | 11-JUL-2000; 2000US-0614150.   |
| XX       |  |
| PA       | (PEKE ) PE CORP NY.  |
| XX       |  |
| PI       | Venter JC, Adams M, Li PWD, Myers EW;  |
| XX       |  |
| DR       | WPI; 2001-656860/75.   |
| DR       | N-PSDB; ABL07662.  |
| XX       |  |
| PT       | New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell   |

PT interactions -  
XX Disclosure; SEQ ID NO 17469; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
CC sequences (ABLI01840-ABLI16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 581 AA;  
  
Query Match 80.4%; Score 37; DB 22; Length 581;  
Best Local Similarity 88.9%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 TESLETLVE 10  
Db 393 TKSLLETLVE 401  
|:|||||  
  
RESULT 7  
AAG38900  
ID AAG38900 standard; Protein; 202 AA.  
AC AAG38900;  
XX  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48057.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0121180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 78.3%; Score 36; DB 21; Length 202;  
 Best Local Similarity 80.0%; Pred.No.59;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||:||||  
 Db 67 TTETLETEVE 76

RESULT 8  
 AAG35722  
 ID AAG35722 standard; Protein; 218 AA.  
 XX  
 AC AAG35722;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 43680.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149436.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.

```
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      78.3%; Score 35; DB 21; Length 218;
Best Local Similarity 80.0%; Pred. No. 64;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTESLETLVE 10
      |||:|||||
Db      67 TTETLETEVE 76

RESULT 9
AAG38899
ID AAG38899 standard; Protein; 272 AA.
AC AAG38899;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48056.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126266.
PR 29-MAR-1999; 99US-0126788.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137503.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
```

```
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0159369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159299.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 78.3%; Score 36; DB 21; Length 272;
Best Local Similarity 80.0%; Pred No. 82;
Matches 8; Conservative 1; Mismatches 0; Gaps 0;
Indels 1;

Qy 1 TTSTLETLVE 10
Db 137 TTSTLETLVE 146

RESULT 10
AAG35721
ID AAG35721 standard; Protein; 288 AA.
AC AAG35721;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 43679.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136382.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
```

PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139809.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 78.3%; Score 36; DB 21; Length 288;  
Best Local Similarity 80.0%; Pred. No. 88;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETVE 10  
|||:||||  
Db 137 TTELETEVE 146

RESULT 11

AAG38898  
ID AAG38898 standard; Protein; 300 AA.  
XX  
AC AAG38898;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48055.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140696.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142330.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.



PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159320.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 78.3%; Score 36; DB 21; Length 300;  
Best Local Similarity 80.0%; Pred. NO. 92;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TTESLETLVE 10  
|||:|:|:|  
Db 165 TTEILETEVE 174

## RESULT 12

AAAG35720  
ID AAG35720 standard; Protein; 316 AA.

XX AAG35720;

AC AAG35720;

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 43678.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 25-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 27-JUN-1999; 99US-0140991.  
PR 28-JUN-1999; 99US-0141287.  
PR 29-JUN-1999; 99US-0141842.  
PR 30-JUN-1999; 99US-0142154.  
PR 01-JUL-1999; 99US-0142055.  
PR 02-JUL-1999; 99US-0142390.  
PR 03-JUL-1999; 99US-0142803.  
PR 04-JUL-1999; 99US-0142920.  
PR 05-JUL-1999; 99US-0142977.  
PR 06-JUL-1999; 99US-0143542.  
PR 07-JUL-1999; 99US-0143624.  
PR 08-JUL-1999; 99US-0144005.  
PR 09-JUL-1999; 99US-0144085.  
PR 10-JUL-1999; 99US-0144086.  
PR 11-JUL-1999; 99US-0144325.  
PR 12-JUL-1999; 99US-0144331.  
PR 13-JUL-1999; 99US-0144332.  
PR 14-JUL-1999; 99US-0144333.  
PR 15-JUL-1999; 99US-0144334.  
PR 16-JUL-1999; 99US-0144335.  
PR 17-JUL-1999; 99US-0144352.  
PR 18-JUL-1999; 99US-0144632.  
PR 19-JUL-1999; 99US-0144684.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 23-JUL-1999; 99US-0145085.  
PR 24-JUL-1999; 99US-0145087.  
PR 25-JUL-1999; 99US-0145089.  
PR 26-JUL-1999; 99US-0145192.  
PR 27-JUL-1999; 99US-0145218.  
PR 28-JUL-1999; 99US-0145224.  
PR 29-JUL-1999; 99US-0145276.  
PR 30-JUL-1999; 99US-0145913.  
PR 31-JUL-1999; 99US-0145918.  
PR 01-AUG-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 03-AUG-1999; 99US-0146386.  
PR 04-AUG-1999; 99US-0146388.  
PR 05-AUG-1999; 99US-0146389.  
PR 06-AUG-1999; 99US-0147038.  
PR 07-AUG-1999; 99US-0147204.  
PR 08-AUG-1999; 99US-0147302.  
PR 09-AUG-1999; 99US-0147192.  
PR 10-AUG-1999; 99US-0147260.  
PR 11-AUG-1999; 99US-0147303.  
PR 12-AUG-1999; 99US-0147416.  
PR 13-AUG-1999; 99US-0147493.  
PR 14-AUG-1999; 99US-0147935.  
PR 15-AUG-1999; 99US-0148171.  
PR 16-AUG-1999; 99US-0148319.  
PR 17-AUG-1999; 99US-0148341.  
PR 18-AUG-1999; 99US-0148565.  
PR 19-AUG-1999; 99US-0148684.  
PR 20-AUG-1999; 99US-0149175.  
PR 21-AUG-1999; 99US-0149226.  
PR 22-AUG-1999; 99US-0149722.  
PR 23-AUG-1999; 99US-0149723.  
PR 24-AUG-1999; 99US-0149929.  
PR 25-AUG-1999; 99US-0149902.  
PR 26-AUG-1999; 99US-0149930.  
PR 27-AUG-1999; 99US-0150566.  
PR 28-AUG-1999; 99US-0150884.  
PR 29-AUG-1999; 99US-0151065.  
PR 30-AUG-1999; 99US-0151066.  
PR 31-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 02-SEP-1999; 99US-0152363.  
PR 03-SEP-1999; 99US-0153070.  
PR 04-SEP-1999; 99US-0153758.  
PR 05-SEP-1999; 99US-0154018.  
PR 06-SEP-1999; 99US-0154039.  
PR 07-SEP-1999; 99US-0154779.  
PR 08-SEP-1999; 99US-0155139.  
PR 09-SEP-1999; 99US-0155486.  
PR 10-SEP-1999; 99US-0155659.  
PR 11-SEP-1999; 99US-0156458.  
PR 12-SEP-1999; 99US-0156596.  
PR 13-SEP-1999; 99US-0157117.  
PR 14-SEP-1999; 99US-0157753.  
PR 15-SEP-1999; 99US-0157865.  
PR 16-SEP-1999; 99US-0158029.  
PR 17-SEP-1999; 99US-0158232.  
PR 18-SEP-1999; 99US-0158369.  
PR 19-SEP-1999; 99US-0159293.  
PR 20-SEP-1999; 99US-0159294.  
PR 21-SEP-1999; 99US-0159295.  
PR 22-SEP-1999; 99US-0159329.  
PR 23-SEP-1999; 99US-0159330.  
PR 24-SEP-1999; 99US-0159331.  
PR 25-SEP-1999; 99US-0159637.  
PR 26-SEP-1999; 99US-0159638.  
PR 27-SEP-1999; 99US-0159584.  
PR 28-SEP-1999; 99US-0160741.  
PR 29-SEP-1999; 99US-0160767.  
PR 30-SEP-1999; 99US-0160768.  
PR 01-OCT-1999; 99US-0160770.  
PR 02-OCT-1999; 99US-0160814.  
PR 03-OCT-1999; 99US-0160815.  
PR 04-OCT-1999; 99US-0160980.  
PR 05-OCT-1999; 99US-0160981.  
PR 06-OCT-1999; 99US-0160989.  
PR 07-OCT-1999; 99US-0161404.  
PR 08-OCT-1999; 99US-0161405.  
PR 09-OCT-1999; 99US-0161406.  
PR 10-OCT-1999; 99US-0161359.  
PR 11-OCT-1999; 99US-0161360.  
PR 12-OCT-1999; 99US-0161361.  
PR 13-OCT-1999; 99US-0161920.  
PR 14-OCT-1999; 99US-0161921.  
PR 15-OCT-1999; 99US-0161992.  
PR 16-OCT-1999; 99US-0161993.  
PR 17-OCT-1999; 99US-0162142.

Query Match 78.3%; Score 36; DB 21; Length 316;  
Best Local Similarity 80.0%; Pred. No. 97;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLVE 10  
|||:|||||  
Db 165 TTETLEVE 174

RESULT 13

AAB76536  
ID AAB76536 standard; Protein; 1277 AA.

XX AAB76536;

XX 11-APR-2001 (first entry)

XX Corynebacterium glutamicum MCT protein SEQ ID NO:54.

XX Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;  
membrane construction and membrane transport protein; petroleum spill;  
hydrocarbon degradation; gram positive aerobic bacterium; marker;  
identification; microorganism; fine chemical production; transformation;  
genome mapping; genetic engineering.

|    |  |  |
|----|--|--|
| XX | Corynebacterium glutamicum.  |  |
| OS |  |  |
| XX | WO200100805-A2.  |  |
| XX | 04-JAN-2001.   |  |
| XX |  |  |
| PD | 23-JUN-2000; 2000WO-IB00926.   |  |
| XX |  |  |
| XX | 25-JUN-1999; 99US-0141031.   |  |
| PR | 08-JUL-1999; 99DE-1031454.   |  |
| PR | 08-JUL-1999; 99DE-1031478.   |  |
| PR | 08-JUL-1999; 99DE-1031563.   |  |
| PR | 09-JUL-1999; 99DE-1032122.   |  |
| PR | 09-JUL-1999; 99DE-1032124.   |  |
| PR | 09-JUL-1999; 99DE-1032125.   |  |
| PR | 09-JUL-1999; 99DE-1032128.   |  |
| PR | 09-JUL-1999; 99DE-1032180.   |  |
| PR | 03-JUL-1999; 99DE-1032182.   |  |
| PR | 09-JUL-1999; 99DE-1032190.   |  |
| PR | 09-JUL-1999; 99DE-1032191.   |  |
| PR | 09-JUL-1999; 99DE-1032209.   |  |
| PR | 09-JUL-1999; 99DE-1032212.   |  |
| PR | 03-JUL-1999; 99DE-1032227.   |  |
| PR | 03-JUL-1999; 99DE-1032228.   |  |
| PR | 09-JUL-1999; 99DE-1032229.   |  |
| PR | 09-JUL-1999; 99DE-1032230.   |  |
| PR | 14-JUL-1999; 99DE-1032927.   |  |
| PR | 14-JUL-1999; 99DE-1033005.   |  |
| PR | 14-JUL-1999; 99DE-1033006.   |  |
| PR | 27-AUG-1999; 99DE-1040764.   |  |
| PR | 27-AUG-1999; 99DE-1040765.   |  |
| PR | 27-AUG-1999; 99DE-1040766.   |  |
| PR | 27-AUG-1999; 99DE-1040830.   |  |
| PR | 27-AUG-1999; 99DE-1040831.   |  |
| PR | 27-AUG-1999; 99DE-1040832.   |  |
| PR | 27-AUG-1999; 99DE-1040833.   |  |
| PR | 31-AUG-1999; 99DE-1041378.   |  |
| PR | 31-AUG-1999; 99DE-1041379.   |  |
| PR | 31-AUG-1999; 99DE-1041395.   |  |
| PR | 03-SEP-1999; 99DE-1042077.   |  |
| PR | 03-SEP-1999; 99DE-1042078.   |  |
| PR | 03-SEP-1999; 99DE-1042079.   |  |
| PR | 03-SEP-1999; 99DE-1042088.   |  |
| XX |  |  |
| XX | (BADI ) BASF AG.   |  |
| XX |  |  |
| PI | Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;              |  |
| XX | WPI; 2001-071486/08.   |  |
| DR | N-PSDB; AAF67769.  |  |
| XX |  |  |
| PT | Corynebacterium glutamicum nucleic acids encoding membrane construction  |  |
| PT | and membrane transport proteins or their portions, useful for typing or  |  |
| PT | identifying C. glutamicum or related bacteria, and as markers for        |  |
| PT | transformation -   |  |
| XX |  |  |
| PS | Claim 20; Page 242-246; 1119pp; English.                                 |  |
| XX |  |  |
| CC | AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane      |  |
| CC | construction and membrane transport (MCT) proteins given in AAB76510 to  |  |
| CC | AAB76847. The MCT nucleic acids and proteins are useful in the           |  |
| CC | identification of microorganisms which can be used to produce fine       |  |
| CC | chemicals, for modulating fine chemical production in C. glutamicum or   |  |
| CC | related bacteria (e.g. Brevibacterium lactofermentum), the typing or     |  |
| CC | identification of C. glutamicum or related bacteria, as reference points |  |
| CC | for mapping C. glutamicum genome, and as markers for transformation.     |  |
| CC | AAF68082 and AAF68082 represent sequencing primers which are used in an  |  |
| CC | example from the present invention.                                      |  |
| XX |  |  |
| SQ | Sequence 1277 AA;  |  |

Query Match 78.3%; Score 36; DB 22; Length 1277;



CC utilised in the genetic engineering of *Corynebacterium glutamicum* and the  
 CC related *Brevibacterium* species and *Acetobacter xylinum* and *Chlorella* to  
 CC make it a better or more efficient producer of one or more fine  
 CC chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs  
 CC having altered activities which indirectly impact the production of one  
 CC or more desired fine chemicals from plants. Primers AAH8705 - AAH80707  
 CC are used in the sequencing of the CMRP cDNA sequences of the invention.  
 XX

SQ Sequence 122 AA;  
 Query Match 73.9%; Score 34; DB 22; Length 122;  
 Best Local Similarity 70.0%; Pred. No. 80;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETLVE 10  
 |||||: |||  
 Db 72 TTESIKKLV 81

RESULT 17  
 AAM00100  
 ID AAM00100 standard; Protein; 305 AA.

XX AAM00100;  
 AC AAM00100;  
 DT 28-SEP-2001 (first entry)

DE Plastidic triosephosphate isomerase sequence #168.

KW Moss; carbohydrate metabolism related protein; CMRP; sugar; cofactor;  
 KW fine chemical production; carbohydrate; polysaccharide.

XX *Physcomitrella patens*.

XX WO200144476-A2.

PN 21-JUN-2001.

PD 14-DEC-2000; 2000WO-EP12697.

PF 16-DEC-1999; 99US-0171101.

XX (BADI ) BASF PLANT SCI GMBH.

XX Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;  
 PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

XX WPI; 2001-398155/42.  
 DR N-PSDB; AAH88786.

XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism  
 PT related protein useful for modulating production of fine chemicals such  
 PT as carbohydrates, cofactors and enzymes from microorganisms and plants

PS Claim 30; Page 128-129; 133pp; English.

XX This invention relates to nucleic acid molecules AAH8708 - AAH8796  
 CC isolated from *Physcomitrella patens* (a moss), which encode carbohydrate  
 CC metabolism related proteins (CMRP) represented in AAM00022 - AAM00110.  
 CC Included in the invention is a vector containing the CMRP cDNA, and a  
 CC host cell transformed with the vector. The host cell (a microorganism,  
 CC *Corynebacterium* or *Brevibacterium*, moss or algae or a plant cell) is  
 CC useful for producing a fine chemical such as carbohydrates, cofactors  
 CC and/or enzymes. The nucleic acid molecules are suitable for modifying a  
 CC carbohydrate production system in a host, e.g., microorganisms and  
 CC plants. They are also useful to identify those DNA sequences and enzymes  
 CC in other species which are useful to modify the biosynthesis of starch,  
 CC cell wall polysaccharides and sugars. The nucleic acid molecules may be  
 CC utilised in the genetic engineering of *Corynebacterium glutamicum* and the  
 CC related *Brevibacterium* species and *Acetobacter xylinum* and *Chlorella* to  
 CC make it a better or more efficient producer of one or more fine  
 CC chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs

CC having altered activities which indirectly impact the production of one  
 CC or more desired fine chemicals from plants. Primers AAH8705 - AAH80707  
 CC are used in the sequencing of the CMRP cDNA sequences of the invention.  
 XX

SQ Sequence 305 AA;  
 Query Match 73.9%; Score 34; DB 22; Length 305;  
 Best Local Similarity 70.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETLVE 10  
 |||||: |||  
 Db 67 TTESIKKLV 76

RESULT 18  
 ABB16242  
 ID ABB16242 standard; Protein; 68 AA.

XX ABB16242;  
 AC ABB16242;

DT 23-JAN-2002 (first entry)

XX Human nervous system related polypeptide SEQ ID NO 4899.

XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antischlicking; antianemic; antiarthritic; cancer;  
 KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX *Homo sapiens*.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

```
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0233080.
PR 08-SEP-2000; 2000US-0233081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249284.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249286.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX N-PSDB; ABA12568.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
XX
XX Claim 11; SEQ ID NO 4899; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (ABA14678-ABA18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 68 AA;
XX
XX Query Match 71.7%; Score 33; DB 22; Length 68;
XX Best Local Similarity 77.8%; Pred. No. 64;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 2 TESLETLVE 10
XX DB 52 TESLHTLVQ 60
```

```

RESULT 19
AAG83035
ID AAG83035 standard; Protein; 153 AA.
XX
AC AAG83035;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:3164.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
DR N-PSDB; AAH53885.
XX
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 835; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 153 AA;
XX
XX Query Match 71.7%; Score 33; DB 22; Length 153;
XX Best Local Similarity 70.0%; Pred. No. 1.6e+02;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TTESLETLVE 10
XX |||||:|
XX 116 TTESIEQAVE 125
XX
XX RESULT 20
XX AAY17249
XX ID AAY17249 standard; Protein; 154 AA.
XX

```

```

AC AAY17249;
XX
XX 09-AUG-1999 (first entry)
XX
XX S. aureus ribH polypeptide.
XX
XX Lumazine synthase; ribH; riboflavin; differential screening; gastritis;
XX H. pylori infection; gastrointestinal carcinoma; gastric ulcer; wound;
XX bacterial adhesion; cell invasion.
XX
XX Staphylococcus aureus.
XX
XX WO9927128-A1.
XX
XX 03-JUN-1999.
XX
XX 24-NOV-1998; 98WO-US25096.
XX
XX 25-NOV-1997; 97US-0977553.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Debouck C, Fedon JC, Jaworski DD, Mooney J, Palmer LM;
XX Traini CM, Wang M, Warren RL, Zhong Y;
XX
XX WPI; 1999-357850/30.
XX
XX N-PSDB; AAX60827.
XX
XX New Staphylococcus aureus ribH polynucleotide and polypeptide useful
XX in the treatment of gastric ulcers and gastritis
XX
XX Claim 14; Page 5; 48pp; English.
XX
XX The invention relates to polynucleotides and polypeptides of the lumazine
XX synthase (beta-subunit) family, designated ribH. They are isolated from
XX Staphylococcus aureus. ribH enzymes are involved in the synthesis of
XX riboflavin. The polynucleotides and polypeptides may be employed as
XX research reagents and material for the discovery of treatments and
XX diagnostics for diseases, particularly human diseases. They can be used
XX for diagnosis of the disease and staging of disease, and as reagents in
XX differential screening methods. The polynucleotides may be used as a
XX source for hybridization probes, and for screening of genetic mutations,
XX serotype, and identification, and for chromosome identification. The
XX polypeptides can be used to produce antibodies. The polypeptides can
XX also be used in vaccine formulations, and to identify agonists and
XX antagonists. These are used to prevent, inhibit or treat diseases,
XX particularly of Helicobacter pylori infections, such as gastrointestinal
XX carcinoma. Gastric ulcers and gastritis may also be treated. The
XX polypeptides can also be used to treat wounds and in-dwelling devices to
XX prevent bacterial adhesion and infection, and to block ribH protein-
XX mediated mammalian cell invasion. Antagonists and agonists of ribH may
XX also be employed to inhibit and treat diseases. The present sequence
XX represents the S. aureus ribH polypeptide.
XX
XX Sequence 154 AA;
XX
XX Query Match 71.7%; Score 33; DB 20; Length 154;
XX Best Local Similarity 70.0%; Pred. No. 1.6e+02;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TTESLETLVE 10
XX |||||:|
XX 116 TTESIEQAVE 125
XX
XX RESULT 21
XX AAB10801
XX ID AAB10801 standard; Protein; 154 AA.
XX
XX AAB10801;
XX
XX 05-FEB-2001 (first entry)
XX

```

DE Staphylococcus aureus lumazine synthase protein.  
 XX Lumazine synthase; capsid; cytostatic; antiviral; antibacterial; vaccine;  
 KW gene therapy; immunotherapy; biosensor; diagnosis.  
 XX Staphylococcus aureus.  
 OS WO200053229-A2.  
 PN 14-SEP-2000.  
 XX 03-MAR-2000; 2000WO-EP01899.  
 XX 08-MAR-1999; 99DE-1010102.  
 XX (FISC/) FISCHER M.  
 PA (BACH/) BACHER A.  
 XX Fischer M, Bacher A;  
 PI WPI; 2000-572230/53.  
 XX Protein conjugate based on lumazine synthase as carrier, useful e.g.  
 PT for vaccination and immunotherapy, contains many functional molecules  
 PT attached to outside of the carrier -  
 XX Disclosure; Figure 8; 180pp; German.  
 PS This invention describes a novel protein conjugate (I) comprising at  
 CC least 1 functional region (FR) at any position in the sequence of a  
 CC carrier protein (II) to form a capsid three-dimensional structure of the  
 CC lumazine synthase (LS) type, such that the outer periphery is covalently  
 CC linked to many FR. The invention also describes (1) a hetero-oligomeric  
 CC protein conjugate (Ia) comprising either a mixture of at least 2  
 CC different (I) or of at least 1 (I) and at least 1 (II) that lacks FR,  
 CC with the components optionally covalently linked by chemical treatment;  
 CC (2) a method for preparation of (I) and (Ia); (3) a vector for  
 CC preparation of (I); (4) a DNA (III) that encodes (I); (5) LS from  
 CC Bacillus subtilis with Cys 93 and/or Cys139 replaced by Ser; (6) DNA  
 CC (IIia) encoding LS from Aquifex aeolicus which is codon-optimized for  
 CC expression in a recombinant Escherichia coli strain; (7) a chimeric  
 CC protein (CP) comprising amino acids (aa) 1-60 from LS of B. subtilis and  
 CC aa 61-154 from LS of A. aeolicus for use as (II); and (8) pharmaceutical  
 CC compositions and vaccines containing (I) and (Ia). The products of the  
 CC invention have cytostatic, antiviral and antibacterial activity and can  
 CC be used in a vaccine or for gene therapy. (I) are used as pharmaceuticals  
 CC (e.g. for immunotherapy of human immune deficiency virus infection or  
 CC tumors) and in vaccines, including multivalent vaccines, against  
 CC bacterial or viral infections, to produce diagnostic or therapeutic  
 CC antibodies, for selective detection, purification and characterization of  
 CC antibodies, and for preparation of protein libraries. (I) may also be  
 CC used as biosensors and for diagnosis, e.g. of tumors. Genes that encode  
 CC (I) are useful in DNA vaccines and for preparation of plant-based oral  
 CC vaccines. (II) can contain many FR (same or different) at the surface of a  
 CC spherical particle (LS comprises 60 subunits that assemble into an  
 CC icosahedron). The large number of FR may increase sensitivity in  
 CC immunosays and the efficiency of immunotherapy agents.  
 XX Sequence 154 AA;  
 SQ  
 Query Match 71.7%; Score 33; DB 21; Length 154;  
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TTESLETLVE 10  
 DB 116 TTESIEQAVE 125  
 RESULT 22  
 ABP40258  
 ID ABP40258 standard; Protein; 157 AA.  
 XX

AC ABP40258;  
 XX 24-JUL-2002 (first entry)  
 XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5103.  
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX Staphylococcus epidermidis.  
 OS US6380370-B1.  
 PN 30-APR-2002.  
 XX 13-AUG-1998; 98US-0134001.  
 XX 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 XX N-PSDB; ABN92803.  
 DR Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PS Disclosure; SEQ ID 5103; 267pp; English.  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP5124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX Sequence 157 AA;  
 SQ  
 Query Match 71.7%; Score 33; DB 23; Length 157;  
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TTESLETLVE 10  
 DB 120 TTESIEQAVE 129  
 RESULT 23  
 AAR27741  
 ID AAR27741 standard; Protein; 501 AA.  
 XX AAR27741;  
 AC AAR27741;  
 XX 25-MAR-2003 (updated)  
 DT 09-MAR-1993 (first entry)  
 XX Sequence transcribed from first reading frame of  
 DE Vaccinia virus DNA from positions 23501-25000.  
 DE Vaccinia virus  
 XX Virus vector; vaccinia virus; papillomavirus; HPV;  
 KW immunotherapeutic; neutral site.  
 XX Vaccinia virus.  
 OS WO9216636-A1.  
 PN



```

XX PD 01-OCT-1992.
XX XX
XX PF 10-MAR-1992; 92WO-CB00424.
XX XX
XX PR 14-MAR-1991; 91GB-0005383.
XX XX
XX PA (IMMU ) IMMUNOLOGY LTD.
XX XX
XX PI Bournsnel MEG, Inglis SC, Munro AJ;
XX XX
XX DR WPI; 1992-349219/42.
XX DR N-PSDB; AAQ29469.
XX XX
XX PT Recombinant virus vectors encoding human papillomavirus proteins
XX PT - for treating and vaccinating against HPV infections and
XX PT conditions caused by them, such as cervical cancer
XX XX
XX PS Disclosure; Fig 21; 83pp; English.
XX XX
XX CC To make a recombinant virus vector comprising human papillomavirus
XX CC genes inserted into the vaccinia virus genome, neutral sites
XX CC for insertion must be utilised such that replicative ability is not
XX CC adversely affected. The neutral sites are identified by analysing
XX CC the viral genome to identify ORFs which are likely to encode
XX CC functional genes and selecting sites between such ORFs or within
XX CC sequences for non-functional genes. The sequence shown is that
XX CC transcribed from the vaccinia virus WR strain positions 23501-25000
XX CC contg. the regions covered by the four fragments salP, G, H and I.
XX CC The sequence was transcribed in all three reading frames to determine
XX CC genuine vaccinia virus genes via codon usage, thus determining neutral
XX CC sites. HPV DNA sequences may be inserted neutral sites, e.g. those
XX CC encoding E6 or E7 of HPV 16 and 18 or mutants of these proteins.
XX CC The recombinant virus vector may be used immunotherapeutically to
XX CC activate cells of the immune system against HPV.
XX CC See also AAR27723-43.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 501 AA;
XX XX
XX Query Match 71.7%; Score 33; DB 13; Length 501;
XX Best Local Similarity 77.8%; Pred. No. 5.9e+02;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX Qy 2 TESLETLVE 10
XX Db 110 TESLERLVE 118
XX
XX RESULT 24
XX ID ABB90301
XX XX
XX AC ABB90301 standard; Protein; 530 AA.
XX XX
XX DT 24-MAY-2002 (first entry)
XX XX
XX DE Human polypeptide SEQ ID NO 2677.
XX XX
XX KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200190304-A2.
XX XX
XX PD 29-NOV-2001.
XX XX
XX PF 18-MAY-2001; 2001WO-US16450.
XX XX

```

```

PR 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Birse CE, Rosen CA;
XX XX
XX DR WPI; 2002-122018/16.
XX DR N-PSDB; ABL90710.
XX XX
XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX PT prevention of neural, immune system, muscular, reproductive,
XX PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX PT disorders -
XX XX
XX PS Claim 11; SEQ ID NO 2677; 2081pp + Sequence Listing; English.
XX XX
XX CC The invention relates to novel genes (ABL9449-ABL90853) and proteins
XX CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating
XX CC medical conditions e.g. by protein or gene therapy. The genes are
XX CC isolated from a range of human tissues disclosed in the specification.
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX CC and parasitic infections.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 530 AA;
XX XX
XX Query Match 71.7%; Score 33; DB 23; Length 530;
XX Best Local Similarity 87.5%; Pred. No. 6.3e+02;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 TTESLETL 8
XX Db 171 STESLETL 178
XX
XX RESULT 25
XX ID AAM92426
XX XX
XX AC AAM92426 standard; Protein; 539 AA.
XX XX
XX DT 06-NOV-2001 (first entry)
XX XX
XX DE Human digestive system antigen SEQ ID NO: 1775.
XX XX
XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX KW digestive system disorder; Meckel's diverticulum.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200155314-A2.
XX XX
XX PD 02-AUG-2001.
XX XX
XX PF 17-JAN-2001; 2001WO-US01324.
XX XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX XX

```

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-502630/55.  
XX N-PSDB; AAK88199.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Claim 11; SEQ ID NO 1775; 986pp; English.  
XX  
CC The present invention provides the protein and coding sequences of a



CC polymorphism, mutation, deletion, truncation and expansion in the gene  
 CC or its gene transcript. Pharmaceutical compositions and gene therapy  
 CC constructs comprising the mucin genes are used for treating disease  
 CC conditions associated with aberrant Mucin expression, altered properties  
 CC of mucus or epithelial inflammatory processes involving Mucins like  
 CC Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and  
 CC colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast  
 CC cancer. The mucin genes and the polypeptides are used for determining  
 CC these diseases or their predisposition. The MUC11 and MUC12 polypeptides  
 CC are used for preparing antagonist and antibodies. The present sequence  
 CC represents the human MUC12 polypeptide.

XX Sequence 585 AA;

SQ Query Match 71.7%; Score 33; DB 21; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 7.1e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
 Db 226 STESLETL 233

RESULT 28  
 ID AM24515 standard; Protein; 585 AA.

XX AC AM24515;

DT 12-OCT-2001 (first entry)

XX C903P predicted amino acid sequence.

DE Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 KW immunogenic; gene therapy; vaccine; colonic cancer.

XX Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 15-FEB-2000; 2000US-0504629.

XX 06-MAR-2000; 2000US-0519444.

XX 19-MAY-2000; 2000US-0575251.

XX 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

XX King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

XX prevention, diagnosis and treatment of colonic cancer -

XX Claim 2; Page 444-445; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity,  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the

CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512  
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences  
 CC given in the exemplification of the present invention.

SQ Sequence 585 AA;

Query Match 71.7%; Score 33; DB 22; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 7.1e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
 Db 226 STESLETL 233

RESULT 29  
 ID ABP55364 standard; Protein; 585 AA.

XX AC ABP55364;

DT 30-JAN-2003 (first entry)

XX Human colon tumour protein for clone C903P SEQ ID NO:1067.

XX Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;  
 KW tumour; immune response; immunostimulant; cytostatic; vaccine.

XX Homo sapiens.

XX WO2000283070-A2.

XX 24-OCT-2002.

XX 09-APR-2002; 2002WO-US11475.

XX 10-APR-2001; 2001US-0833263.

XX 03-AUG-2001; 2001US-0922217.

XX 19-DEC-2001; 2001US-0025380.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

XX Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD;

XX Skeiky YAW, Fanger GR, Vedvick TS, Carter D;

XX WPI; 2003-067548/06.

XX N-PSDB; ABZ33689.

XX New polynucleotide, useful for the preparation of a composition for

XX stimulating an immune response against, or treating, cancer -

XX Disclosure; Page 445-447; 537pp; English.

XX The present invention describes compounds (I) for the immunotherapy and  
 CC diagnosis of colon cancer. Also described: (1) a method for detecting  
 CC the presence of cancer in a patient; (2) a method for stimulating and/or  
 CC expanding T cells specific for a tumour protein; (3) an isolated T cell  
 CC population comprising T cells prepared by the method of (2); (4) a method  
 CC for stimulating an immune response in a patient; (5) a method for  
 CC treating cancer in a patient; and (6) a method for inhibiting the

CC development of cancer in a patient. (I) have immunostimulant and  
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725  
CC and ABP5343 to ABP5391 represent human colon cancer/tumour related  
CC sequences used in the exemplification of the present invention.

XX SQ Sequence 585 AA;

Query Match 71.7%; Score 33; DB 24; Length 585;  
Best Local Similarity 87.5%; Pred. No. 7.1e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||

DB 226 STESLETL 233

RESULT 30  
ABB97588  
ID ABB97588 standard; Protein; 601 AA.

XX AC ABB97588;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 856.

XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;  
KW antiferility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag.

OS Homo sapiens.

XX WO200222660-A2.

PN 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI; 2002-292408/33.

DR N-PSDB; ABN32774.

XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -

XX Claim 20; SEQ ID NO 856; 509pp; English.

XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention.

XX SQ Sequence 601 AA;

Query Match 71.7%; Score 33; DB 23; Length 601;  
Best Local Similarity 87.5%; Pred. No. 7.3e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||

DB 220 STESLETL 227

RESULT 31

AM79256  
ID AAM79256 standard; Protein; 615 AA.

XX AC AAM79256;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1918.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;

DR WPI; 2001-476283/51.

DR N-PSDB; AAK52389.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

XX Claim 20; Page 4316-4318; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX SQ Sequence 615 AA;

Query Match 71.7%; Score 33; DB 22; Length 615;  
Best Local Similarity 87.5%; Pred. No. 7.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
:|||||

DB 256 STESLETL 263

Thu Oct 30 14:52:59 2003

us-09-868-293b-2.rag

```

RESULT 32
AAW20828
ID AAW20828 standard; Protein; 2440 AA.
XX
AC AAW20828;
XX
DT 16-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein 11gsl0309orf15.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR ) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgaard BL;
XX
DR WPI; 1997-052306/05.
DR N-PSDB; AAT68081.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 61; Page 1233-1238; 1481pp; English.
XX
CC The present sequence shows a Helicobacter pylori cytoplasmic protein
CC that may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 2440 AA;
XX
Query Match 71.7%; Score 33; DB 18; Length 2440;
Best Local Similarity 60.0%; Pred. No. 3.5e+03;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 TTESLETLVE 10
Db 1141 TTQSLETLTK 1150
RESULT 33
ABBO3560
ID ABBO3560 standard; Protein; 92 AA.
XX
AC ABBO3560;
XX
DT 08-JAN-2002 (first entry)
XX

```

```

XX DE Human musculoskeletal system related polypeptide SEQ ID NO 1507.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein;
XX KW musculoskeletal system.
XX OS Homo sapiens.
XX PN WO200155367-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01338.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 14-AUG-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225759.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.

```

PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244674.  
 PR 08-NOV-2000; 2000US-0244675.  
 PR 08-NOV-2000; 2000US-0244676.  
 PR 08-NOV-2000; 2000US-0244677.  
 PR 08-NOV-2000; 2000US-0244678.  
 PR 08-NOV-2000; 2000US-0244679.  
 PR 08-NOV-2000; 2000US-0245523.  
 PR 08-NOV-2000; 2000US-0245524.  
 PR 08-NOV-2000; 2000US-0245525.  
 PR 08-NOV-2000; 2000US-0245526.  
 PR 08-NOV-2000; 2000US-0245527.  
 PR 08-NOV-2000; 2000US-0245528.  
 PR 08-NOV-2000; 2000US-0245529.  
 PR 08-NOV-2000; 2000US-0245532.  
 PR 08-NOV-2000; 2000US-0245533.  
 PR 08-NOV-2000; 2000US-0245603.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0255719.

PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-451937/48.  
 DR N-PSDB; AAL35142.  
 XX  
 PT Isolated polypeptide for treating, preventing and/ or prognosing  
 PT disorders related to the musculoskeletal system including  
 PT musculoskeletal cancers and also for testing and detection e.g.  
 PT diagnosis -  
 XX  
 XX Claim 11; SEQ ID NO 1507; 781pp + Sequence Listing; English.  
 PS  
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins  
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful  
 CC for preventing, treating or ameliorating medical conditions e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids, proteins,  
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
 CC other cancers of the adrenal gland, bone, bone marrow, breast,  
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound  
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
 CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 92 AA;  
 Query Match 69.6%; Score 32; DB 22; Length 92;  
 Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TTESLETLVE 10  
 Db 54 TTESLETLVE 63  
 RESULT 34  
 ABU12854  
 ID ABU12854 standard; Protein; 92 AA.  
 XX  
 AC ABU12854;  
 XX  
 DT 26-FEB-2003 (first entry)  
 XX  
 DE Novel human musculoskeletal system antigen #474.  
 XX  
 KW Musculoskeletal system antigen; cancer; metastasis;  
 KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
 KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
 KW post-operative tissue repair; limb regeneration; neuronal growth;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW AIDS-related complex; chondrocyte growth; bone regeneration;  
 KW periodontal regeneration; tissue transport; bone graft; skin aging;  
 KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
 KW cell growth; organ transplant; cell differentiation; body height;

weight; hair colour; eye colour; skin; percentage of adipose tissue;  
pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;  
depression; tendency for violence; pain; reproductive capability;  
hormone level; endocrine level; appetite; libido; memory; stress;  
storage capability; fat content; lipid content; protein content;  
carbohydrate content; vitamin content; cofactor content;  
nutritional component.

XX Homo sapiens.  
OS US200214740-A1.  
XX 10-OCT-2002.  
XX 17-JAN-2001; 2001US-0764877.  
XX 31-JAN-2000; 2000US-179065P.  
XX 04-FEB-2000; 2000US-180628P.  
XX 28-JUN-2000; 2000US-214886P.  
XX 07-JUL-2000; 2000US-216647P.  
XX 07-JUL-2000; 2000US-216880P.  
XX 11-JUL-2000; 2000US-217487P.  
XX 11-JUL-2000; 2000US-217496P.  
XX 14-JUL-2000; 2000US-218290P.  
XX 26-JUL-2000; 2000US-220963P.  
XX 14-AUG-2000; 2000US-224518P.  
XX 14-AUG-2000; 2000US-224519P.  
XX 14-AUG-2000; 2000US-225267P.  
XX 14-AUG-2000; 2000US-225268P.  
XX 14-AUG-2000; 2000US-225270P.  
XX 14-AUG-2000; 2000US-225447P.  
XX 14-AUG-2000; 2000US-225757P.  
XX 14-AUG-2000; 2000US-225758P.  
XX 22-AUG-2000; 2000US-226868P.  
XX 30-AUG-2000; 2000US-228924P.  
XX 01-SEP-2000; 2000US-229287P.  
XX 01-SEP-2000; 2000US-229343P.  
XX 01-SEP-2000; 2000US-229344P.  
XX 01-SEP-2000; 2000US-229345P.  
XX 05-SEP-2000; 2000US-229509P.  
XX 05-SEP-2000; 2000US-229513P.  
XX 08-SEP-2000; 2000US-231143P.  
XX 21-SEP-2000; 2000US-234223P.  
XX 21-SEP-2000; 2000US-234274P.  
XX 25-SEP-2000; 2000US-234997P.  
XX 27-SEP-2000; 2000US-235834P.  
XX 29-SEP-2000; 2000US-236327P.  
XX 29-SEP-2000; 2000US-236367P.  
XX 29-SEP-2000; 2000US-236368P.  
XX 29-SEP-2000; 2000US-236369P.  
XX 29-SEP-2000; 2000US-236370P.  
XX 02-OCT-2000; 2000US-236802P.  
XX 02-OCT-2000; 2000US-237037P.  
XX 02-OCT-2000; 2000US-237038P.  
XX 02-OCT-2000; 2000US-237039P.  
XX 13-OCT-2000; 2000US-237040P.  
XX 20-OCT-2000; 2000US-239935P.  
XX 20-OCT-2000; 2000US-240960P.  
XX 20-OCT-2000; 2000US-241785P.  
XX 20-OCT-2000; 2000US-241809P.  
XX 01-NOV-2000; 2000US-244617P.  
XX 17-NOV-2000; 2000US-249299P.  
XX 08-DEC-2000; 2000US-251856P.  
XX 08-DEC-2000; 2000US-251868P.  
XX 08-DEC-2000; 2000US-251869P.

(ROSE/) ROSEN C A.  
(RUBE/) RUBEN S M.  
(BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

DR WPI; 2003-128199/12.  
DR N-PSDB; ABX58130.  
XX Isolated nucleic acid molecules encoding musculoskeletal system  
XX associated polypeptides, useful for detecting disorders, e.g. cancer -  
PT  
PT  
XX Claim 11; SEQ ID NO 1507; 31pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule comprising a  
XX sequence encoding musculoskeletal system associated polypeptides useful  
XX for detecting disorders, e.g., cancer or cancer metastases, in animals  
XX or humans. The nucleic acid: stimulates re-vascularisation of ischaemic  
XX tissues associated with conditions such as thrombosis, arteriosclerosis,  
XX and other cardiovascular conditions; treats wounds due to injuries,  
XX burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
XX and limb regeneration; stimulates neuronal growth; can treat and prevent  
XX neuronal damage occurring in certain disorders or neurodegenerative  
XX conditions, such as, Alzheimer's disease, Parkinson's disease, and  
XX AIDS-related complex; stimulates chondrocyte growth, thus they can be  
XX used to enhance bone and periodontal regeneration and aid in tissue  
XX transports or bone grafts; prevents skin aging due to sunburn by  
XX stimulating keratinocyte growth; prevents hair loss, since FGF family  
XX members activate hair-forming cells and promotes melanocyte growth;  
XX stimulates growth and differentiation of hematopoietic cells and bone  
XX marrow cells when used in combination with other cytokines; maintains  
XX organs before transplantation or for supporting cell culture of primary  
XX tissues; induces tissue of mesodermal origin to differentiate in early  
XX embryos; increases or decreases the differentiation or proliferation of  
XX embryonic stem cells, besides, haematopoietic lineage; modulates  
XX mammalian characteristics, such as, body height, weight, hair colour, eye  
XX colour, skin, percentage of adipose tissue, pigmentation, size, and shape  
XX (e.g., cosmetic surgery); modulates mammalian metabolism; changes  
XX mammal's metal state or physical state by influencing biorhythms,  
XX cardiac rhythms, depression, tendency for violence, tolerance for pain,  
XX reproductive capabilities, hormonal or endocrine levels, appetite,  
XX libido, memory, or stress; increases or decreases storage capabilities,  
XX fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors  
XX or other nutritional components. This is the amino acid sequence of a  
XX novel human musculoskeletal system antigen.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from the US patent office at  
XX ftp.segdata.uspto.gov/sequence.html?DocID=20020147140.  
XX  
XX Sequence 92 AA;

Query Match 69.6%; Score 32; DB 24; Length 92;  
Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
DB 54 TTESIEQAIE 63

RESULT 35  
AAW28030  
ID AAW28030 standard; Protein; 125 AA.

XX AAW28030;  
XX 27-AUG-1998 (first entry)

DE Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
KW Staphylococcal gene; regulatory element; bacterial gene expression;  
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
XX toxic shock syndrome.

OS Staphylococcus aureus.

XX WO9730070-A1.



XX PD 21-AUG-1997.  
 XX PF 19-FEB-1997; 97WO-US02318.  
 XX PR 20-FEB-1996; 96US-0011888.  
 XX PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
 XX PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;  
 XX DR WPI; 1997-424969/39.  
 XX DR N-PSDB; AAT83984.  
 XX PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used  
 PT to isolate antimicrobial compounds, and in vaccines against S.  
 PT aureus infection  
 XX PS Claim 6; Page 419; 989pp; English.  
 XX CC The present sequence represents a Staphylococcus aureus protein of  
 CC unknown function. The DNA sequence was isolated from a library of  
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can  
 CC be used in the construction of ribozymes and antisense sequences to  
 CC control the expression of Staphylococcal genes. The DNA sequence is  
 CC also useful as a source of regulatory elements for the control of  
 CC bacterial gene expression. The present protein may be used to produce  
 CC vaccines to enable a host to produce specific antibodies with  
 CC antibacterial action. These vaccines and antibodies would protect  
 CC a host against invasion by S. aureus, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled  
 CC skin syndrome, and toxic shock syndrome.  
 XX SQ Sequence 125 AA;  
 Query Match 69.6%; Score 32; DB 18; Length 125;  
 Best Local Similarity 77.8%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 TTESLETLV 9  
 ||:|||||  
 Db 101 TTKSLEALV 109  
 RESULT 36  
 AAB10806  
 ID AAB10806 standard; Protein; 156 AA.  
 XX AC AAB10806;  
 XX DT 05-FEB-2001 (first entry)  
 XX DE Shigella flexneri lumazine synthase protein.  
 XX KW Lumazine synthase; capsid; cytostatic; antiviral; antibacterial; vaccine;  
 KW gene therapy; immunotherapy; biosensor; diagnosis.  
 XX OS Shigella flexneri.  
 XX PN WO200053229-A2.  
 XX PD 14-SEP-2000.  
 XX PF 03-MAR-2000; 2000WO-EP01899.  
 XX PR 08-MAR-1999; 99DE-1010102.  
 XX (FISC/) FISCHER M.  
 PA (BACH/) BACHER A.  
 PI Fischer M, Bacher A;  
 XX PA

DR WPI; 2000-572230/53.  
 XX Protein conjugate based on lumazine synthase as carrier, useful e.g.  
 PT for vaccination and immunotherapy, contains many functional molecules  
 PT attached to outside of the carrier -  
 XX Disclosure; Figure 8; 180pp; German.  
 XX CC This invention describes a novel protein conjugate (I) comprising at  
 CC least 1 functional region (FR) at any position in the sequence of a  
 CC carrier protein (II) to form a capsid three-dimensional structure of the  
 CC lumazine synthase (LS) type, such that the outer periphery is covalently  
 CC linked to many FR. The invention also describes (I) a hetero-oligomeric  
 CC protein conjugate (Ia) comprising either a mixture of at least 2  
 CC different (I) or of at least 1 (I) and at least 1 (II) that lacks FR,  
 CC with the components optionally covalently linked by chemical treatment;  
 CC (2) a method for preparation of (I) and (Ia); (3) a vector for  
 CC preparation of (I); (4) a DNA (III) that encodes (I); (5) LS from  
 CC Bacillus subtilis with Cys 93 and/or Cys139 replaced by Ser; (6) DNA  
 CC (IIia) encoding LS from Aquifex aeolicus which is codon-optimized for  
 CC expression in a recombinant Escherichia coli strain; (7) a chimeric  
 CC protein (CP) comprising amino acids (aa) 1-60 from LS of B. subtilis and  
 CC aa 61-154 from LS of A. aeolicus for use as (II); and (8) pharmaceutical  
 CC compositions and vaccines containing (I) and (Ia). The products of the  
 CC invention have cytostatic, antiviral and antibacterial activity and can  
 CC be used in a vaccine or for gene therapy. (I) are used as pharmaceuticals  
 CC (e.g. for immunotherapy of human immune deficiency virus infection or  
 CC tumors) and in vaccines, including multivalent vaccines, against  
 CC bacterial or viral infections, to produce diagnostic or therapeutic  
 CC antibodies, for selective detection, purification and characterization of  
 CC antibodies, and for preparation of protein libraries. (I) may also be  
 CC used as biosensors and for diagnosis, e.g. of tumors. Genes that encode  
 CC (I) are useful in DNA vaccines and for preparation of plant-based oral  
 CC vaccines. (I) can contain many FR (same or different) at the surface of a  
 CC spherical particle (LS comprises 60 subunits that assemble into an  
 CC icosahedron). The large number of FR may increase sensitivity in  
 CC immunoassays and the efficiency of immunotherapy agents.  
 XX Sequence 156 AA;  
 Query Match 69.6%; Score 32; DB 21; Length 156;  
 Best Local Similarity 60.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TTESLETLVE 10  
 ||:|||||  
 Db 118 TTSEIQAIE 127  
 RESULT 37  
 AAB10807  
 ID AAB10807 standard; Protein; 156 AA.  
 XX AC AAB10807;  
 XX DT 05-FEB-2001 (first entry)  
 XX DE Escherichia coli lumazine synthase protein.  
 XX KW Lumazine synthase; capsid; cytostatic; antiviral; antibacterial; vaccine;  
 KW gene therapy; immunotherapy; biosensor; diagnosis.  
 XX OS Escherichia coli.  
 XX PN WO200053229-A2.  
 XX PD 14-SEP-2000.  
 XX PF 03-MAR-2000; 2000WO-EP01899.  
 XX PR 08-MAR-1999; 99DE-1010102.  
 XX (FISC/) FISCHER M.  
 PA

Thu Oct 30 14:52:59 2003

PA (BACH/) BACHER A.  
 XX Fischer M, Bacher A;  
 PI  
 XX  
 XX  
 DR WPI; 2000-572230/53.  
 XX  
 XX Protein conjugate based on lumazine synthase as carrier, useful e.g.  
 PT for vaccination and immunotherapy, contains many functional molecules  
 PT attached to outside of the carrier  
 XX  
 XX Disclosure; Figure 8; 180pp; German.  
 XX  
 XX This invention describes a novel protein conjugate (I) comprising at  
 CC least 1 functional region (FR) at any position in the sequence of a  
 CC carrier protein (II) to form a capsid three-dimensional structure of the  
 CC lumazine synthase (LS) type, such that the outer periphery is covalently  
 CC linked to many FR. The invention also describes (I) a hetero-oligomeric  
 CC protein conjugate (Ia) comprising either a mixture of at least 2  
 CC different (I) or of at least 1 (I) and at least 1 (II) that lacks FR,  
 CC with the components optionally covalently linked by chemical treatment;  
 CC (2) a method for preparation of (I) and (Ia); (3) a vector for  
 CC preparation of (I); (4) a DNA (III) that encodes (I); (5) LS from  
 CC Bacillus subtilis with Cys 93 and/or Cys139 replaced by Ser; (6) DNA  
 CC (IIa) encoding LS from Aquifex aeolicus which is codon-optimized for  
 CC expression in a recombinant Escherichia coli strain; (7) a chimeric  
 CC protein (Cp) comprising amino acids (aa) 1-60 from LS of B. subtilis and  
 CC aa 61-154 from LS of A. aeolicus for use as (II); and (8) pharmaceutical  
 CC compositions and vaccines containing (I) and (Ia). The products of the  
 CC invention have cytostatic, antiviral and antibacterial activity and can  
 CC be used in a vaccine or for gene therapy. (I) are used as pharmaceuticals  
 CC (e.g. for immunotherapy of human immune deficiency virus infection or  
 CC tumors) and in vaccines, including multivalent vaccines, against  
 CC bacterial or viral infections, to produce diagnostic or therapeutic  
 CC antibodies, for selective detection, purification and characterization of  
 CC antibodies, and for preparation of protein libraries. (I) may also be  
 CC used as biosensors and for diagnosis, e.g. of tumors. Genes that encode  
 CC vaccines. (I) can contain many FR (same or different) at the surface of a  
 CC spherical particle (LS comprises 60 subunits that assemble into an  
 CC icosahedron). The large number of FR may increase sensitivity in  
 CC immunoassays and the efficiency of immunotherapy agents.  
 XX  
 XX Sequence 156 AA;  
 XX  
 XX Query Match 69.6%; Score 32; DB 21; Length 156;  
 XX Best Local Similarity 60.0%; Pred. No. 2.5e+02;  
 XX Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 XX QY 1 TTESLETLVE 10  
 XX |||||:|:  
 XX Db 118 TTESIEQAIE 127  
 XX  
 XX RESULT 38  
 XX AAB70801  
 XX ID AAB70801 standard; Protein; 156 AA.  
 XX  
 XX AC AAB70801;  
 XX  
 XX DT 08-JUN-2001 (first entry)  
 XX  
 XX DE E. coli ribE protein SEQ ID 1.  
 XX  
 XX KW 6,7-dimethyl-8-ribityl-lumazine synthase; Flavin biosynthesis pathway;  
 XX inhibitor; plant; herbicide; ribE.  
 XX  
 XX OS Escherichia coli.  
 XX  
 XX PN DE19942175-A1.  
 XX  
 XX PD 08-MAR-2001.  
 XX  
 XX PF 03-SEP-1999; 99DE-1042175.  
 XX PA

XX 03-SEP-1999; 99DE-1042175.  
 XX (BACH/) BACHER A.  
 XX  
 XX Bacher A, Moertl S;  
 PI  
 XX  
 XX WPI; 2001-258919/27.  
 DR  
 XX  
 XX New isolated 6,7-dimethyl-8-ribityl-lumazine synthase protein, useful  
 PT for screening for 6,7-dimethyl-8-ribityl-lumazine synthase inhibitors  
 PT potentially useful as herbicides  
 XX  
 XX Disclosure; Page 15-16; 22pp; German.  
 XX  
 XX This invention describes a novel isolated 6,7-dimethyl-8-ribityl-lumazine  
 CC synthase protein (I). The invention also describes (1) isolated DNA  
 CC encoding (I), and optionally at least one other enzyme of the flavin  
 CC biosynthesis pathway; (2) a screening assay for 6,7-dimethyl-8-ribityl-  
 CC lumazine synthase inhibitors, comprising incubating an aqueous mixture  
 CC containing (I), 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinone (II)  
 CC and 3,4-dihydroxy-2-butanone 4-phosphate (III) in the presence and  
 CC absence of a test compound; (3) a screening assay for inhibition-resistant  
 CC ribityl-lumazine formed; (3) a screening assay for inhibition-resistant  
 CC 6,7-dimethyl-8-ribityl-lumazine synthase mutants, comprising incubating  
 CC an aqueous mixture containing a mutant of (I), (II) and (III) in the  
 CC presence and absence of a specific 6,7-dimethyl-8-ribityl-lumazine  
 CC synthase inhibitor, and comparing the amounts of 6,7-dimethyl-8-ribityl-  
 CC lumazine formed; and (4) inhibiting 6,7-dimethyl-8-ribityl-lumazine  
 CC synthase in a plant, by treatment with a compound identified by the assay  
 CC of (2). (I) is useful for screening for 6,7-dimethyl-8-ribityl-lumazine  
 CC synthase inhibitors, which are potentially useful as herbicides. This  
 CC sequence represents the E. coli ribE protein used to illustrate the  
 CC method of the invention.  
 XX  
 XX Sequence 156 AA;  
 XX  
 XX Query Match 69.6%; Score 32; DB 22; Length 156;  
 XX Best Local Similarity 60.0%; Pred. No. 2.5e+02;  
 XX Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 XX QY 1 TTESLETLVE 10  
 XX |||||:|:  
 XX Db 118 TTESIEQAIE 127  
 XX  
 XX RESULT 39  
 XX AAO05824  
 XX ID AAO05824 standard; Protein; 164 AA.  
 XX  
 XX AC AAO05824;  
 XX  
 XX DT 06-NOV-2001 (first entry)  
 XX  
 XX DE Human polypeptide SEQ ID NO 19716.  
 XX  
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX nervous system disorders; arthritis; inflammation.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX PN WO200164835-A2.  
 XX  
 XX PD 07-SEP-2001.  
 XX  
 XX PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 XX PR 28-FEB-2000; 2000US-0515126.  
 XX  
 XX PR 18-MAY-2000; 2000US-0577409.  
 XX  
 XX PA (HYSE-) HYSEQ INC.

```

XX PI Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-514838/56.
XX N-PSDB; AAI85755.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX Claim 20; SEQ ID NO 19716; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 164 AA;

Query Match 69.6%; Score 32; DB 22; Length 164;
Best Local Similarity 66.7%; Pred. NO. 2.6e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTESLETLV 9
Db 113 TTESLKSLLI 121

RESULT 40
AAP50078
ID AAP50078 standard; Protein; 169 AA.
XX AC AAP50078;
XX DT 25-MAR-2003 (updated)
XX DT 30-OCT-1991 (first entry)
XX DE Sequence of the C-terminal of the mitochondrial uncoupling
XX DE protein of rat brown adipose tissue.
XX KW Obesity therapy; brown adipose tissue.
XX OS Rattus norvegicus.
XX PN EP148686-A.
XX PD 17-JUL-1985.
XX PF 18-DEC-1984; 84EP-0402627.
XX PR 21-DEC-1983; 83FR-0020487.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX PA (INSP ) INST PASTEUR & GENETIC SYSTEMS.
XX PI Ricquier D, Boullaud F, Thibault J, Weissenbach J;
XX WPI: 1985-173055/29.
XX N-PSDB; AAN50090.
XX DNA probes for detecting brown adipose tissue gene - contg.
XX PT nucleic acid sequence recognising uncoupling protein gene or mRNA
XX
PS Disclosure; Fig 2; 27pp; French.
XX The inventors claim probes for detecting the gene or mRNA coding for
CC synthesis of the uncoupling protein of mammalian brown adipose tissue
CC mitochondria. The probes may be used as research tools for screening
CC drugs for the ability to induce brown adipose tissue (BAT) formation,
CC e.g. for treatment of obesity. The probes are produced by isolating
CC the 17S fraction of the BAT polyA RNA from cold-adapted animals
CC and using it to prepare cDNA upon which the probes are based.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 169 AA;

Query Match 69.6%; Score 32; DB 6; Length 169;
Best Local Similarity 87.5%; Pred. NO. 2.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETL 8
Db 28 TTESLSTL 35

Search completed: October 30, 2003, 14:16:14
Job time : 85 secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2003, 14:14:22 ; Search time 39 Seconds

(without alignments)  
24.659 Million cell updates/sec

Title: US-09-868-293B-2

Perfect score: 46

Sequence: 1 TTESLETLVE 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 46    | 100.0       | 129    | 2 H86500 | L7/L12 ribosomal p |
| 2          | 46    | 100.0       | 129    | 2 C72122 | ribosomal protein  |
| 3          | 46    | 100.0       | 130    | 2 A71530 | ribosomal protein  |
| 4          | 46    | 100.0       | 130    | 2 H81684 | ribosomal protein  |
| 5          | 36    | 78.3        | 316    | 2 C96733 | hypothetical prote |
| 6          | 35    | 76.1        | 734    | 2 F88098 | protein F18A12.4 [ |
| 7          | 35    | 76.1        | 966    | 2 D96662 | hypothetical prote |
| 8          | 34    | 73.9        | 282    | 2 S61860 | hrpC protein - Pse |
| 9          | 34    | 73.9        | 401    | 2 D82692 | 8-amino-7-oxonon   |
| 10         | 34    | 73.9        | 438    | 2 B82077 | outer membrane pro |
| 11         | 34    | 73.9        | 508    | 2 T07971 | aromatic-L-amino-a |
| 12         | 34    | 73.9        | 728    | 2 T24716 | hypothetical prote |
| 13         | 34    | 73.9        | 895    | 2 T32374 | hypothetical prote |
| 14         | 33    | 71.7        | 124    | 2 B42526 | B3R protein - vacc |
| 15         | 33    | 71.7        | 154    | 2 A89962 | 6,7-dimethyl-8-rib |
| 16         | 33    | 71.7        | 155    | 2 H72857 | AcOrf-63 protein - |
| 17         | 33    | 71.7        | 167    | 2 J01797 | B3R protein - vacc |
| 18         | 33    | 71.7        | 179    | 2 T37441 | probable 20.9K pro |
| 19         | 33    | 71.7        | 206    | 2 AD1898 | urase accessory p  |
| 20         | 33    | 71.7        | 257    | 2 C87550 | polysaccharide dea |
| 21         | 33    | 71.7        | 773    | 2 T44989 | transducer protein |
| 22         | 33    | 71.7        | 780    | 2 T50315 | hypothetical prote |
| 23         | 33    | 71.7        | 784    | 2 D84386 | ATP-dependent RNA  |
| 24         | 33    | 71.7        | 231    | 2 D71870 | hypothetical prote |
| 25         | 33    | 71.7        | 386    | 1 GNWVDF | genome polyprotein |
| 26         | 33    | 71.7        | 3824   | 2 S37431 | ankyrin 2, neuro   |
| 27         | 33    | 71.7        | 4447   | 2 A69679 | polyketide synthas |
| 28         | 32    | 69.6        | 156    | 2 S26202 | riboflavin synthas |
| 29         | 32    | 69.6        | 156    | 2 D90687 | riboflavin synthas |

|    |    |      |     |          |                    |
|----|----|------|-----|----------|--------------------|
| 30 | 32 | 69.6 | 156 | 2 AF0386 | riboflavin synthas |
| 31 | 32 | 69.6 | 156 | 2 H85537 | riboflavin synthas |
| 32 | 32 | 69.6 | 156 | 2 A10553 | 6,7-dimethyl-8-rib |
| 33 | 32 | 69.6 | 159 | 2 F84091 | hypothetical prote |
| 34 | 32 | 69.6 | 216 | 2 D89827 | conserved hypotet  |
| 35 | 32 | 69.6 | 306 | 2 A31106 | mitochondrial unco |
| 36 | 32 | 69.6 | 307 | 2 A26294 | uncoupling protein |
| 37 | 32 | 69.6 | 373 | 2 F84396 | signal recognition |
| 38 | 32 | 69.6 | 391 | 2 AE0370 | probable sulfatase |
| 39 | 32 | 69.6 | 406 | 2 D64647 | conserved hypotet  |
| 40 | 32 | 69.6 | 409 | 2 G71936 | hypothetical prote |
| 41 | 32 | 69.6 | 422 | 2 C71930 | udp-n-acetylmuramo |
| 42 | 32 | 69.6 | 466 | 2 AF0224 | flagellar hook-ass |
| 43 | 32 | 69.6 | 481 | 2 T48516 | probable oligopept |
| 44 | 32 | 69.6 | 598 | 2 T38241 | probable U3 small  |
| 45 | 32 | 69.6 | 718 | 2 A82352 | iron(III) compound |

## ALIGNMENTS

### RESULT 1

H86500

L7/L12 ribosomal protein [imported] - Chlamydomophila pneumoniae (strain J138)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: H86500

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; I

Nucleic Acids Res. 28, 2311-2314, 2000

A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: H86500

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <STO>

A:Cross-references: GB:BA000008; NID:g8978453; PIDN:BAA98290.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: r17

C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 100.0%; Score 46; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 0.064; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10

Db 2 TTESLETLVE 11

### RESULT 2

C72122

ribosomal protein L7/L12 CP0695 [imported] - Chlamydomophila pneumoniae (strains CWL029 a

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: C72122; G81548

R:Kalmann, S.; Mitchell, W.; Marathe, R.; Lamm, C.; Fan, J.; Olinger, L.; Grimwood, J.

Nature Genet. 21, 385-389, 1999

A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: C72122

A:Molecule type: DNA

A:Residues: 1-129 <ARN>

A:Cross-references: GB:AE001593; GB:AE001363; NID:g4376334; PIDN:AA18233.1; PID:g437633

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg

Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: G81548

A:Molecule type: DNA

A:Residues: 1-129 <REA>

A;Cross-references: GB:AE002228; GB:AE002161; NID:g7189606; PIDN:AAF38503.1; PID:g7189606  
 A;Experimental source: strain AR39, HL cells  
 C;Genetics:  
 A;Gene: r17; CP0695  
 C;Superfamily: Escherichia coli ribosomal protein L12

Query Match 100.0%; Score 46; DB 2; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 0.064; 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0

QY 1 TTSETLVE 10  
 Db 2 TTSETLVE 11

RESULT 3  
 A;Cross-references: GB:AE001304; GB:AE001273; NID:g3328730; PIDN:AAC67909.1; PID:g3328730  
 A;Experimental source: serotype D, strain UW-3/Cx  
 R;Engel, J.N.; Pollack, J.; Malik, F.; Ganem, D.  
 J. Bacteriol. 172, 5732-5741, 1990  
 A;Title: Cloning and characterization of RNA polymerase core subunits of Chlamydia trachomatis  
 A;Reference number: A36135; MUID:91008945; PMID:2211507  
 A;Accession: B36135  
 A;Molecule type: DNA  
 A;Residues: 1-130 <ARN>  
 A;Cross-references: GB:AE001304; GB:AE001273; NID:g3328730; PIDN:AAC67909.1; PID:g3328730  
 A;Experimental source: serotype D, strain UW-3/Cx  
 R;Engel, J.N.; Pollack, J.; Malik, F.; Ganem, D.  
 J. Bacteriol. 172, 5732-5741, 1990  
 A;Title: Cloning and characterization of RNA polymerase core subunits of Chlamydia trachomatis  
 A;Reference number: A36135; MUID:91008945; PMID:2211507  
 A;Accession: B36135  
 A;Molecule type: DNA  
 A;Residues: 1-130 <ENG>  
 A;Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A;Molecule type: DNA  
 C;Genetics:  
 A;Gene: r17  
 C;Superfamily: Escherichia coli ribosomal protein L12  
 C;Keywords: acetylated amino end; methylated amino acid; protein biosynthesis; ribosome  
 F;2-130/Product: ribosomal protein L7/L12 #status predicted <Mar>  
 F;2/Modified site: acetylated amino end (Thr) (in mature form) #status predicted  
 F;78,89/Modified site: N6-methyllysine (Lys) #status predicted

Query Match 100.0%; Score 46; DB 2; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTSETLVE 10  
 Db 2 TTSETLVE 11

RESULT 4  
 H81684  
 ribosomal protein L7/L12 TC0590 [imported] - Chlamydia muridarum (strain Nigg)  
 C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
 C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
 C;Accession: H81684  
 R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
 A;Reference number: A81500; MUID:20150255; PMID:10684935  
 A;Accession: H81684  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-130 <TT>  
 A;Cross-references: GB:AE002328; GB:AE002160; NID:g7190627; PIDN:AAF39422.1; PID:g7190627  
 A;Experimental source: strain Nigg (MoPn)  
 C;Genetics:

A;Gene: TC0590  
 C;Superfamily: Escherichia coli ribosomal protein L12

Query Match 100.0%; Score 46; DB 2; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTSETLVE 10  
 Db 2 TTSETLVE 11

RESULT 5  
 C96733  
 hypothetical protein FL5H11.10 [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C;Accession: C96733  
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: C96733  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-316 <STO>  
 A;Cross-references: GB:AE005173; NID:g5902401; PIDN:AAD55503.1; GSPDB:GN00141  
 C;Genetics:  
 A;Gene: FL5H11.10  
 A;Map position: 1

Query Match 78.3%; Score 36; DB 2; Length 316;  
 Best Local Similarity 80.0%; Pred. No. 18;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLVE 10  
 Db 165 TTSETLVE 174

RESULT 6  
 F88098  
 protein FL8A12.4 [imported] - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C;Accession: F88098  
 R;anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo  
 A;Reference number: A75000; MUID:99069613; PMID:9851916  
 A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_e  
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an  
 A;Accession: F88098  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-734 <STO>  
 A;Cross-references: GB:chr\_II; PIDN:AB66076.1; PID:g2315626; GSPDB:GN00020; CESP:F18A12  
 C;Genetics:  
 A;Gene: FL8A12.4  
 A;Map position: 2

Query Match 76.1%; Score 35; DB 2; Length 734;  
 Best Local Similarity 66.7%; Pred. No. 74;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTSETLVE 10

Db 364 TESLDTHIE 372  
||||:|:|

## RESULT 7

D9662  
hypothetical protein F24D7.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96662  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D96662  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-966 <STO>  
A:Cross-references: GB:AE005173; NID:g6456160; PIDN:AAF09148.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F24D7.8  
A:Map position: 1

Query Match 76.1%; Score 35; DB 2; Length 966;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETLV 9  
|:|||||

Db 809 TCESLETLV 817  
|:|||||

## RESULT 8

S61860  
hrpC protein - Pseudomonas solanacearum  
C:Species: Pseudomonas solanacearum  
C>Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Oct-1999  
C:Accession: S61860  
R:van Gisegeim, F.; Gough, C.; Zischek, C.; Niqueux, E.; Arlat, M.; Genin, S.; Barberis,  
Mol. Microbiol. 15, 1095-1114, 1995  
A:Title: The hrp gene locus of Pseudomonas solanacearum, which controls the production of  
is complex.  
A:Reference number: S61846; MUID:95349395; PMID:7623665  
A:Accession: S61860  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-282 <VAN>  
A:Cross-references: EMBL:Z14056; NID:g5503397; PIDN:CAA78437.1; PID:g550415  
C:Genetics:  
A:Gene: hrpC

Query Match 73.9%; Score 34; DB 2; Length 282;  
Best Local Similarity 77.8%; Pred. No. 40;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTESLETLV 9  
|:|||||

Db 14 TSLESLETL 22  
|:|||||

## RESULT 9

D92692  
8-amino-7-oxononanoate synthase XF1357 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: D82692

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequel  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:2035717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82692  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-401 <SIM>  
A:Cross-references: GB:AE003967; GB:AE003849; NID:g9106347; PIDN:AAF84166.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; f  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, f  
as-Neto, E.; Docena, C.; El-Doroty, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.S.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, f  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaes  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1357

C:Superfamily: 5-aminolevulinatase synthase; glycine C-acetyltransferase homology  
C:Keywords: phosphoprotein; pyridoxal phosphate  
F:243/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 73.9%; Score 34; DB 2; Length 401;  
Best Local Similarity 60.0%; Pred. No. 60;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETLV 10  
|:|||||

Db 374 TTEQIDTLVQ 383  
|:|||||

## RESULT 10

B82077  
outer membrane protein TolC VC2436 [imported] - Vibrio cholerae (strain N16961 serogrou  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: B82077  
R:Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, f  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20408833; PMID:10952301

A:Accession: B82077

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <HEI>

A:Cross-references: GB:AE004313; GB:AE003852; NID:g9657007; PIDN:AAF95579.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2436

A:Map position: 1

C:Superfamily: outer membrane protein tolC

Query Match 73.9%; Score 34; DB 2; Length 438;  
Best Local Similarity 70.0%; Pred. No. 66;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTESLETLV 10  
|:|||||

Db 226 TTESSEALIE 235  
|:|||||

## RESULT 11

Thu Oct 30 14:53:00 2003

T07971  
aromatic-L-amino-acid decarboxylase (EC 4.1.1.28) 8 - opium poppy  
C:Species: Papaver somniferum (opium poppy)  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 31-Mar-2000  
C:Accession: T07971  
R:Facchini, P.J.; Penzes-Yost, C.; Samanani, N.; Kowalchuk, B.  
Plant Physiol. 118, 69-81, 1998  
A:Title: Expression patterns conferred by tyrosine/dihydroxyphenylalanine decarboxylase  
C:Superfamily: aromatic-L-amino-acid decarboxylase; animal histidine decarboxylase homol  
A:Reference number: Z16257; MUID:98404235; PMID:9733527  
A:Accession: T07971  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-508 <FAC>  
A:Cross-references: EMBL:AF025432; NID:g3282520; PIDN:AAC61841.1; PID:g3282521  
A:Experimental source: strain Marianne  
C:Genetics:  
A:Gene: tydc8  
C:Function:  
A:Description: catalyzes the decarboxylation of aromatic amino acids, including tryptoph  
C:Superfamily: aromatic-L-amino-acid decarboxylase; animal histidine decarboxylase homol  
C:Keywords: carbon-carbon lyase; carboxy-lyase  
F:21-496/Domain: animal histidine decarboxylase homology <HDC>  
  
Query Match 73.9%; Score 34; DB 2; Length 508;  
Best Local Similarity 66.7%; Pred. No. 78;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 TTESLTLVE 10  
|||:||||:  
Db 69 TESLTLVE 77  
  
RESULT 12  
T24716  
hypothetical protein T08G5.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24716  
R:Smye, R.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19927  
A:Accession: T24716  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-728 <WIL>  
A:Cross-references: EMBL:Z81589; PIDN:CAB04724.1; GSPDB:GN00023; CESP:T08G5.2  
C:Genetics:  
A:Gene: CESP:T08G5.2  
A:Map position: 5  
A:Introns: 23/3; 84/3; 119/3; 181/2; 261/1; 304/2; 366/1; 462/3; 523/3; 569/1; 659/3  
  
Query Match 73.9%; Score 34; DB 2; Length 728;  
Best Local Similarity 70.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 TTESLTLVE 10  
|:||||:|  
Db 416 TCESLTLVLQ 425  
  
RESULT 13  
T32374  
hypothetical protein K10F12.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 28-Jul-2000  
C:Accession: T32374  
R:Wohlmann, P.; Beck, C.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid K10F12.  
A:Reference number: Z21157  
A:Accession: T32374  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-895 <WOH>  
A:Cross-references: EMBL:AF025462; PIDN:AAB71005.1; GSPDB:GN00021; CESP:K10F12.3  
A:Experimental source: strain Bristol N2; clone K10F12  
C:Genetics:  
A:Gene: CESP:K10F12.3  
A:Map position: 3  
A:Introns: 46/1; 96/1; 124/3; 165/1; 223/3; 282/3; 340/3; 381/1; 485/3; 516/2; 549/1; 6;  
C:Superfamily: yeast 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase; 1-phosph  
osphodiesterase domain Y homology  
F:338-487/Domain: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain X ho  
  
Query Match 73.9%; Score 34; DB 2; Length 895;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 TTESLTLVE 10  
|||:||||:  
Db 287 TTDNCETLVE 296  
  
RESULT 14  
B42526  
B3 protein - vaccinia virus (strain Copenhagen)  
C:Species: vaccinia virus  
A:Note: host Homo sapiens (man)  
C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 08-Apr-1994  
C:Accession: B42526  
R:Johnson, G.P.  
submitted to GenBank, June 1990  
A:Reference number: A33172  
A:Accession: B42526  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <JOH>  
  
Query Match 71.7%; Score 33; DB 2; Length 124;  
Best Local Similarity 77.8%; Pred. No. 25;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 TESLTLVE 10  
|||:||||:  
Db 94 TESLTLVE 102  
  
RESULT 15  
A89962  
6,7-dimethyl-8-ribityllumazine synthase [imported] - Staphylococcus aureus (strain N315  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: A89962  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: A89962  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-154 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3701561; PIDN:BA842854.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: ribH  
C:Superfamily: riboflavin synthase beta chain  
  
Query Match 71.7%; Score 33; DB 2; Length 154;  
Best Local Similarity 70.0%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 TTESLTLVE 10  
|||:||||:  
Db 94 TESLTLVE 102



Db 116 TTESIQAIVE 125

RESULT 16

H72857

AcOrf-63 protein - Autographa californica nuclear polyhedrosis virus

C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV

A:Note: dsDNA virus

C:Date: 12-Nov-1999

C:Accession: H72857

R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.

Virology 202, 586-605, 1994

A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.

A:Reference number: A72850; MUID:94303173; PMID:8030224

A:Accession: H72857

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <AYR>

A:Cross-references: GB:L22858; NID:9510708; PIDN:AAA66693.1; PID:9559132

C:Genetics:

A:Gene: AcOrf-63

Query Match 71.7%; Score 33; DB 2; Length 155;

Best Local Similarity 70.0%; Pred. No. 32;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTSETLTVLVE 10

|||||

Db 56 TRESLNTLVD 65

RESULT 17

JQ1797

B3R protein - vaccinia virus (strain WR)

C:Species: vaccinia virus

C:Date: 30-Sep-1993

C:Accession: JQ1797

R:Smith, G.L.; Chan, Y.S.; Howard, S.T.

J. Gen. Virol. 72, 1349-1376, 1991

A:Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right in

A:Reference number: JQ1767; MUID:91259063; PMID:2045793

A:Accession: JQ1797

A:Molecule type: DNA

A:Residues: 1-167 <SMI>

A:Cross-references: DDBJ:D11079; NID:g222717; PIDN:BAA01833.1; PID:d1002309; PID:g222748

Query Match 71.7%; Score 33; DB 2; Length 167;

Best Local Similarity 77.8%; Pred. No. 35;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TBSLETLVE 10

|||||

Db 94 TESIERLVE 102

RESULT 18

T37441

Probable 20.9K protein - vaccinia virus (strain Ankara)

C:Species: vaccinia virus

A:Variety: strain Ankara

C:Date: 21-Jan-2000

C:Accession: T37441

R:Antoine, G.; Scheifflinger, F.; Falkner, F.G.; Dörner, F.

submitted to the EMBL Data Library, March 1997

A:Description: the complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain

A:Reference number: Z20877

A:Accession: T37441

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-179 <ANT>

A:Cross-references: EMBL:U94848; PIDN:AAB96546.1

A:Experimental source: strain Ankara

C:Genetics:

A:Note: MVA170R

Query Match 71.7%; Score 33; DB 2; Length 179;

Best Local Similarity 77.8%; Pred. No. 38;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TESLETLVE 10

|||||

Db 34 TESIERLVE 42

RESULT 19

AD1898

urease accessory protein F [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001

C:Accession: AD1898

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD1898

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA000019; PID:gl7130079; GSPDB:GNO0179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr0734

C:Superfamily: Klebsiella urease accessory protein uref

Query Match 71.7%; Score 33; DB 2; Length 206;

Best Local Similarity 77.8%; Pred. No. 45;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TESLETLVE 10

|||||

Db 7 SEGLETLVE 15

RESULT 20

C87550

polyaccharide deacetylase [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001

C:Accession: C87550

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. n. J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: AB7249; MUID:21173698; PMID:11259647

A:Accession: C87550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-257 <STO>

A:Cross-references: GB:AE005673; NID:gl3423971; PIDN:AAK24399.1; GSPDB:GNO0148

C:Genetics:

A:Gene: CC2428

Query Match 71.7%; Score 33; DB 2; Length 257;

Best Local Similarity 60.0%; Pred. No. 57;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTSETLTVLVE 10

|||||

Db 227 TTEALERLID 236

RESULT 21

T44989

transducer protein trII [imported] - Haloarcula vallismortis  
 C:Species: Haloarcula vallismortis  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 01-Dec-2000  
 C:Accession: T44989; S55296; S49233  
 R:Rodwald, K.; Seidel, R.; Engelhardt, M.; Oesterheld, D.  
 submitted to the EMBL Data Library, September 1999  
 A:Description: Primary structure of vHtrII, a transducer protein from Haloarcula vallismortis  
 A:Reference number: Z22889  
 A:Accession: T44989  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-773 <ROD>  
 A:Cross-references: EMBL:AJ249640; PIDN:CAB56464.1  
 A:Experimental source: strain SPI  
 R:Seidel, R.; Scharf, B.; Gautel, M.; Kleine, K.; Oesterheld, D.; Engelhardt, M.  
 Proc. Natl. Acad. Sci. U.S.A. 92, 3036-3040, 1995  
 A:Title: The primary structure of sensory rhodopsin II: a member of an additional retina  
 A:Reference number: S55296; MUID:95224074; PMID:7708770  
 A:Accession: S55296  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 341-773 <SEI>  
 A:Cross-references: EMBL:Z35308; NID:G516321; PIDN:CAA84549.1; PID:G516322  
 C:Genetics:  
 A:Gene: trII; htrII  
 C:Function:  
 A:Description: transducer protein for sensory rhodopsin II  
 C:Superfamily: Halobacterium salinarum transducer protein htrII  
 C:Keywords: signal transduction; transmembrane protein

Query Match 71.7%; Score 33; DB 2; Length 773;  
 Best Local Similarity 60.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||:|||||  
 Db 663 TVDALETLVE 672

RESULT 22  
 T50315  
 hypothetical protein SPBC1703.02 with ARID DNA-binding domain [imported] - fission yeast  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
 C:Accession: T50315  
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Galibert, F.  
 submitted to the EMBL Data Library, January 2000  
 A:Reference number: Z25061  
 A:Accession: T50315  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-780 <MCD>  
 A:Cross-references: EMBL:AL136536; PIDN:CAB66446.1; GSPDB:GN00067; SPDB:SPBC1703.02  
 A:Experimental source: strain 972h(-); cosmid c1703  
 C:Genetics:  
 A:Gene: SPDB.SPBC1703.02  
 A:Map position: 2  
 A:Introns: 38/2

Query Match 71.7%; Score 33; DB 2; Length 780;  
 Best Local Similarity 87.5%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10  
 :|||||  
 Db 695 DSLETLVE 702

RESULT 23  
 D84386  
 ATP-dependent RNA helicase homolog eIF-4A [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 10-May-2001

C:Accession: D84386  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laaky, J.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl, Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L.  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: D84386  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-784 <STO>  
 A:Cross-references: GB:AE004437; NID:G10581766; PIDN:AAG20456.1; GSPDB:GN00138  
 C:Genetics:  
 C:Superfamily: ATP-dependent RNA helicase eIF-4A

Query Match 71.7%; Score 33; DB 2; Length 784;  
 Best Local Similarity 60.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||:|||||  
 Db 689 TTETETIAE 698

RESULT 24  
 D71870  
 hypothetical protein jhp0928 - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C:Accession: D71870  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: D71870  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2231 <ARN>  
 A:Cross-references: GB:AE001522; GB:AE001439; NID:G4155505; PIDN:AAD06506.1; PID:G4155505  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: jhp0928

Query Match 71.7%; Score 33; DB 2; Length 2231;  
 Best Local Similarity 60.0%; Pred. No. 6.6e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||:|||||  
 Db 1085 TTQSLETTIK 1094

RESULT 25  
 GNWVDF  
 genome polyprotein - dengue virus type 4  
 N:Contains: capsid protein; envelope protein; membrane protein; nonstructural protein 5, nonstructural protein NS4a; nonstructural protein NS4b  
 C:Species: dengue virus type 4  
 C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 19-Jan-2001  
 C:Accession: A94352; A94364; A26897; A29121  
 R:Zhao, B.; Mackow, E.; Buckler-White, A.; Markoff, L.; Chanock, R.M.; Lai, C.J.; Makino, S.; Virolgy 155, 77-88, 1986  
 A:Title: Cloning full-length dengue type 4 viral DNA sequences: analysis of genes coding for structural and nonstructural proteins  
 A:Reference number: A94352; MUID:87044106; PMID:3022479  
 A:Accession: A94352  
 A:Molecule type: genomic RNA  
 A:Residues: 1-776 <ZHA>  
 A:Cross-references: GB:M14931  
 R:Mackow, E.; Makino, Y.; Zhao, B.; Zhang, Y.M.; Markoff, L.; Buckler-White, A.; Guiler, A.

Virology 159, 217-228, 1987

A:Title: The nucleotide sequence of dengue type 4 virus: analysis of genes coding for non-structural proteins  
A:Reference number: A94364; MUID:87293881; PMID:3039728

A:Accession: A94364

A:Molecule type: genomic RNA

A:Residues: 774-3386 <NAC>

A:Cross-references: GB:M17255

C:Superfamily: yellow fever virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;

F:2-113/Product: capsid protein #status predicted <CAP>

F:42-58/Domain: transmembrane #status predicted <TM1>

F:100-116/Domain: transmembrane #status predicted <TM2>

F:114-279/Product: membrane protein precursor #status predicted <MEP>

F:114-204/Domain: nonterminal signal sequence #status predicted <SIG>

F:205-279/Product: membrane protein #status predicted <MEM>

F:267-283/Domain: transmembrane #status predicted <TM3>

F:280-773/Product: envelope protein #status predicted <ENV>

F:728-744/Domain: transmembrane #status predicted <TM4>

F:753-769/Domain: transmembrane #status predicted <TM5>

F:774-1184/Product: nonstructural protein NS1 #status predicted <NS1>

F:1157-1179/Domain: transmembrane #status predicted <TM6>

F:1185-1343/Product: nonstructural protein NS2a #status predicted <N2A>

F:1344-1473/Product: nonstructural protein NS2b #status predicted <N2B>

F:1474-2091/Product: nonstructural protein NS3 #status predicted <NS3>

F:1666-1673/Region: nucleotide-binding motif A (P-loop)

F:1753-1758/Region: nucleotide-binding motif B

F:1757-1760/Region: DEAH motif

F:2092-2374/Product: nonstructural protein NS4a #status predicted <N4A>

F:2375-2486/Product: nonstructural protein NS4b #status predicted <N4B>

F:2487-3386/Product: nonstructural protein NS5 #status predicted <NS5>

F:182.346,432,750,903,980,2296,2300,2341,2382,2452,2582,2639,2699,2719,2913,3310/Binding

Query Match 71.7%; Score 33; DB 1; Length 3386;

Best Local Similarity 87.5%; Pred. No. 1.1e+03; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TESLETLV 9

Db 2140 TESLETLM 2147

|||||

RESULT 26

S37431

N:Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid

N:Contains: ankyrin 2, short form

C:Species: Homo sapiens (man)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 13-Aug-1999

C:Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569

R:Chan, W.

submitted to the EMBL Data Library, September 1993

A:Reference number: S37431

A:Accession: S37431

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-3924 <CHA>

A:Cross-references: EMBL:226634; NID:g406287; PIDN:CAA81387.1; PID:g406288

R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.

J. Cell Biol. 114, 241-253, 1991

A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a

A:Reference number: A39643; MUID:91302466; PMID:1830053

A:Accession: A39643

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2077 <OT1>

A:Cross-references: GB:X56957

A:Accession: B39643

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1443,3585-3924 <OTT>

A:Cross-references: EMBL:X56958

R:Tse, W.T.; Menninger, J.C.; Yang-Peng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,

Genomics 10, 858-866, 1991

A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.  
A:Reference number: A40334; MUID:92009921; PMID:1833308

A:Accession: A40334

A:Molecule type: DNA

A:Residues: 463-474, 'PE', 477-495 <TSE>

A:Cross-references: GB:M37123; NID:g178647; PIDN:AAAG2828.1; PID:g178648

R:Chan, W.; Kordeli, E.; Bennett, V.

J. Cell Biol. 123, 1463-1473, 1993

A:Title: 440-kD ankyrinB: structure of the major developmentally regulated domain and s.

A:Reference number: A49462; MUID:94075409; PMID:8253844

A:Accession: A49462

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-3924 <RES>

A:Cross-references: EMBL:226634; NID:g406287; PIDN:CAA81387.1; PID:g406288

C:Genetics:

A:Gene: GDB:ANK2

A:Cross-references: GDB:127607; OMIM:106410

A:Map position: 4q25-4q27

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing

F:2-3924/Product: ankyrin 2, long form #status predicted <WAT>

F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <WAT>

F:63-95/Domain: ankyrin repeat homology <AN01>

F:96-128/Domain: ankyrin repeat homology <AN02>

F:129-161/Domain: ankyrin repeat homology <AN03>

F:162-190/Domain: ankyrin repeat homology <AN04>

F:191-223/Domain: ankyrin repeat homology <AN05>

F:232-264/Domain: ankyrin repeat homology <AN06>

F:265-297/Domain: ankyrin repeat homology <AN07>

F:298-330/Domain: ankyrin repeat homology <AN08>

F:331-363/Domain: ankyrin repeat homology <AN09>

F:364-396/Domain: ankyrin repeat homology <AN10>

F:397-429/Domain: ankyrin repeat homology <AN11>

F:430-462/Domain: ankyrin repeat homology <AN12>

F:463-495/Domain: ankyrin repeat homology <AN13>

F:496-528/Domain: ankyrin repeat homology <AN14>

F:529-561/Domain: ankyrin repeat homology <AN15>

F:562-594/Domain: ankyrin repeat homology <AN16>

F:595-627/Domain: ankyrin repeat homology <AN17>

F:628-660/Domain: ankyrin repeat homology <AN18>

F:661-693/Domain: ankyrin repeat homology <AN19>

F:694-726/Domain: ankyrin repeat homology <AN20>

F:727-759/Domain: ankyrin repeat homology <AN21>

F:760-792/Domain: ankyrin repeat homology <AN22>

F:793-825/Domain: ankyrin repeat homology <AN23>

Query Match 71.7%; Score 33; DB 2; Length 3924;

Best Local Similarity 75.0%; Pred. No. 1.2e+03;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 ESLETLVE 10

Db 3508 ESLETLIE 3515

|||||

RESULT 27

A69679

polyketide synthase pksK - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 03-Nov-2000

C:Accession: A69679

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber-

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch-

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle-

ich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue-

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero-



C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Dec-2001  
 C:Accession: H85537  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: H85537  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-156 <STO>  
 A:Cross-references: GB:AE005174; NID:gl2513270; PIDN:AA054764.1; GSPDB:GN00145; UMGF:Z05  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: Z0516  
 C:Superfamily: riboflavin synthase beta chain  
 C:Keywords: transferase

Query Match 69.6%; Score 32; DB 2; Length 156;  
 Best Local Similarity 60.0%; Pred. No. 52;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||||:|  
 Db 118 TTESIEQAIE 127

RESULT 32  
 A:0553  
 6,7-dimethyl-8-ribityllumazine synthase (riboflavin synthase beta chain) [imported] - Sa  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: A10553  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connerston, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: A10553  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-156 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD08873.1; PID:gl6501685; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0456  
 C:Superfamily: riboflavin synthase beta chain

Query Match 69.6%; Score 32; DB 2; Length 156;  
 Best Local Similarity 60.0%; Pred. No. 52;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||||:|  
 Db 118 TTESIEQAIE 127

RESULT 33  
 F84091  
 hypothetical protein BH3534 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F84091  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: F84091  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:gi0176109; PIDN:BAB07253.1; GSPDB:GN0  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH3534

Query Match 69.6%; Score 32; DB 2; Length 159;  
 Best Local Similarity 70.0%; Pred. No. 53;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTESLETLVE 10  
 |||||:|  
 Db 148 TTTELKLLVE 157

RESULT 34  
 D89827  
 conserved hypothetical protein SA0543 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: D89827  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oga  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: D89827  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-216 <KUR>  
 A:Cross-references: GB:BA000018; PID:gl3700478; PIDN:BAB41775.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA0543

Query Match 69.6%; Score 32; DB 2; Length 216;  
 Best Local Similarity 77.8%; Pred. No. 75;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETLV 9  
 |||||:|  
 Db 192 TTKSLEALV 200

RESULT 35  
 A31106  
 mitochondrial uncoupling protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 21-Jul-2000  
 C:Accession: A31106  
 R:Kozak, L.P.; Britton, J.H.; Kozak, U.C.; Wells, J.M.  
 J. Biol. Chem. 263, 12274-12277, 1988  
 A:Title: The mitochondrial uncoupling protein gene. Correlation of exon structure to tr.  
 A:Reference number: A31106; MUID:88315014; PMID:3410843  
 A:Accession: A31106  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1-306 <KOZ>  
 A:Cross-references: GB:U634418; NID:gl519064; PIDN:AAB07367.1; PID:gl519065  
 C:Genetics:  
 A:Gene: Ucp  
 A:Introns: 41/3; 108/1; 175/1; 209/1; 269/2  
 C:Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology  
 C:Keywords: duplication; mitochondrion; transmembrane protein  
 F:9-102/Domain: ADP,ATP carrier protein repeat homology <ACPL>  
 F:109-201/Domain: ADP,ATP carrier protein repeat homology <ACPD>  
 F:208-295/Domain: ADP,ATP carrier protein repeat homology <ACPD>

Query Match 69.6%; Score 32; DB 2; Length 306;  
 Best Local Similarity 87.5%; Pred. No. 11e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTESLETL 8  
 |||||:|

Thu Oct 30 14:53:00 2003

Db 165 TTESLSTL 172

RESULT 36  
A26294  
N;Alternative names: UCP  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 18-Feb-2000  
C;Accession: A26294 #sequence\_revision 22-Jul-1987 #text\_change 18-Feb-2000  
R;Bouillaud, F.; Weissenbach, J.; Ricquier, D.  
J. Biol. Chem. 261, 1487-1490, 1986  
A;Reference number: A26294; MUID:86111804; PMID:3753702  
A;Accession: A26294  
A;Molecule type: mRNA  
A;Residues: 1-307 <BOU>  
A;Cross-references: GB:M1814; NID:9207556; PIDN:AAA19671.1; PID:9207557  
R;Ridley, R.G.; Patel, H.V.; Gerber, G.E.; Morton, R.C.; Freeman, K.B.  
Nucleic Acids Res. 14, 4025-4035, 1986  
A;Title: Complete nucleotide and derived amino acid sequence of cDNA encoding the mitoch  
A;Reference number: A29278; MUID:86232540; PMID:3012461  
A;Accession: A29278  
A;Molecule type: mRNA  
A;Residues: 1-307 <RI2>  
A;Cross-references: GB:X03894; GB:M15500; NID:957446; PIDN:CAA27531.1; PID:957447  
R;Bouillaud, F.; Raibault, S.; Ricquier, D.  
Biochem. Biophys. Res. Commun. 157, 783-792, 1988  
A;Title: The gene for rat uncoupling protein: complete sequence, structure of primary tr  
A;Reference number: S03842; MUID:89076317; PMID:3202878  
A;Accession: S03842  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-307 <BO2>  
A;Cross-references: EMBL:X12925; NID:957444; PIDN:CAA31392.1; PID:957445  
R;Ridley, R.G.; Patel, H.V.; Parfett, C.L.J.; Olynky, K.A.; Reichling, S.; Freeman, K.B.  
Biosci. Rep. 6, 87-94, 1986  
A;Title: Immunological detection of cDNA clones encoding the uncoupling protein of brown  
A;Reference number: A61566; MUID:86188126; PMID:2421800  
A;Accession: A61566  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 254-307 <RID>  
C;Comment: The source of this protein was brown adipocyte mitochondria.  
C;Genetics:  
A;Introns: 42/3; 109/1; 176/1; 210/1; 270/2  
C;Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology  
C;Keywords: duplication; mitochondrion; transmembrane protein  
F;10-103/Domain: ADP,ATP carrier protein repeat homology <ACP1>  
F;110-202/Domain: ADP,ATP carrier protein repeat homology <ACP2>  
F;209-296/Domain: ADP,ATP carrier protein repeat homology <ACP3>

Query Match 69.6%; Score 32; DB 2; Length 307;  
Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETL 8  
Db 166 TTESLSTL 173

RESULT 37  
F84396  
signal recognition particle receptor [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: F84396  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Leibauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl  
Jung, K.H.; Alam, M.; Freitas, F.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: F84396  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-373 <STO>  
A;Cross-references: GB:AE004437; NID:g10581862; PIDN:AAG20538.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: dpa  
C;Superfamily: cell division protein ftsy

Query Match 69.6%; Score 32; DB 2; Length 373;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTESLETLVE 10  
Db 138 TESTDTLIE 146

RESULT 38  
AE0370  
probable sulfatase modifier protein YPO3046 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AE0370  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B  
;geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AE0370  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-391 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC92288.1; PID:g15980999; GSPDB:GN00175  
C;Genetics:  
A;Gene: YPO3046  
C;Superfamily: arylsulfatase activating enzyme atSB

Query Match 69.6%; Score 32; DB 2; Length 391;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTESLETLVE 10  
Db 35 TESTLETLPIE 44

RESULT 39  
D64647  
conserved hypothetical protein HP1020 - Helicobacter pylori (strain 26695)  
C;Species: Helicobacter pylori  
C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 07-Mar-2003  
C;Accession: D64647  
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khatkhat, H.G.; Glodek, A.; McKen  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L  
Nature 388, 539-547, 1997  
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C  
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A;Reference number: A64520; MUID:97394467; PMID:9252185  
A;Accession: D64647  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-406 <TOM>  
A;Cross-references: GB:AE000610; GB:AE000511; NID:g2314160; PIDN:AA08064.1; PID:g23141  
C;Superfamily: bifunctional 4-diphosphocytidyl-2-methyl-D-erythritol synthase/2C-methyl

Query Match 69.6%; Score 32; DB 2; Length 406;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTESLETLVE 10

```

Db      |.|.|.|.|.:
25 TLETLETLIK 34

RESULT 40
G71936
hypothetical protein jhp0404 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 07-Mar-2003
C:Accession: G71936
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: G71936
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-409 <ARN>
A:Cross-references: GB:AE001474; GB:AE001439; NID:G4154929; PIDN:AAD05981.1; PID:G415493
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0404
C:Superfamily: bifunctional 4-diphosphocytidyl-2-methyl-D-erythritol synthase/2C-methyl-
Query Match      69.6%; Score 32; DB 2; Length 409;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTSETLVE 10
        |.|.|.|.:
Db      28 TLETLETLIK 37

Search completed: October 30, 2003, 14:19:17
Job time : 42 secs

```

**This Page Blank (uspto)**